

Deposit & Copying of Dissertation Declaration



UNIVERSITY OF
CAMBRIDGE

Board of Graduate Studies

Please note that you will also need to bind a copy of this Declaration into your final, hardbound copy of thesis - this has to be the very first page of the hardbound thesis.

1	Surname (Family Name)	Forenames(s)	Title
	LEPOUTRE	MAXIME , CHARLES	MR
2	Title of Dissertation as approved by the Degree Committee		
	DEMOCRATIC SPEECH IN DIVIDED TIMES		

In accordance with the University Regulations in *Statutes and Ordinances* for the PhD, MSc and MLitt Degrees, I agree to deposit one print copy of my dissertation entitled above and one print copy of the summary with the Secretary of the Board of Graduate Studies who shall deposit the dissertation and summary in the University Library under the following terms and conditions:

1. Dissertation Author Declaration

I am the author of this dissertation and hereby give the University the right to make my dissertation available in print form as described in 2. below.

My dissertation is my original work and a product of my own research endeavours and includes nothing which is the outcome of work done in collaboration with others except as declared in the Preface and specified in the text. I hereby assert my moral right to be identified as the author of the dissertation.

The deposit and dissemination of my dissertation by the University does not constitute a breach of any other agreement, publishing or otherwise, including any confidentiality or publication restriction provisions in sponsorship or collaboration agreements governing my research or work at the University or elsewhere.

2. Access to Dissertation

I understand that one print copy of my dissertation will be deposited in the University Library for archival and preservation purposes, and that, unless upon my application restricted access to my dissertation for a specified period of time has been granted by the Board of Graduate Studies prior to this deposit, the dissertation will be made available by the University Library for consultation by readers in accordance with University Library Regulations and copies of my dissertation may be provided to readers in accordance with applicable legislation.

3	Signature	Date
		03/10/2018

Corresponding Regulation

Before being admitted to a degree, a student shall deposit with the Secretary of the Board one copy of his or her hard-bound dissertation and one copy of the summary (bearing student's name and thesis title), both the dissertation and the summary in a form approved by the Board. The Secretary shall deposit the copy of the dissertation together with the copy of the summary in the University Library where, subject to restricted access to the dissertation for a specified period of time having been granted by the Board of Graduate Studies, they shall be made available for consultation by readers in accordance with University Library Regulations and copies of the dissertation provided to readers in accordance with applicable legislation.

DEMOCRATIC SPEECH IN DIVIDED TIMES

Maxime Charles Lepoutre
Trinity College, University of Cambridge
Date of Submission: April 2018

This dissertation is submitted for the degree of Doctor of Philosophy.

DEMOCRATIC SPEECH IN DIVIDED TIMES

Maxime Charles Lepoutre

Abstract

Democratic theorists have influentially argued that inclusive deliberation, where citizens voice their concerns and exchange justifications, is crucial to democracy. However, this deliberative ideal has come under sustained attack for being excessively utopian. As a result, to make this ideal more relevant and action-guiding, the present thesis investigates what norms should govern deliberation in political settings marked by severe social divisions.

After motivating this project (Chapter 1), I defend the following account of deliberation. Although the requirement that deliberators appeal to shared reasons is morally attractive, even the weakest variant of this norm risks excluding too many considerations from the public deliberation of divided societies. To offset these exclusionary tendencies, I argue that public deliberation should give a greater role to emotionally-charged forms of speech, such as narrative (Chapter 2). Now, this last suggestion might seem overly *inclusive*, by opening the door for intensely angry narratives and for narratives expressing degrading or disrespectful views. In response to this ‘overinclusiveness’ concern, I argue 1) that degrading or disrespectful public speech is best countered through state-backed counterspeech, rather than through coercive legal norms that forcibly eliminate it from public discourse (Chapter 3) and 2) that narratives expressing anger in fact have a crucial epistemic role to play in divided societies, by enhancing our understanding of persisting injustices (Chapter 4).

The final two chapters address a pressing worry: that the deliberative norms I advance demand too much of actual citizens. One might think that if citizens distrust each other and are highly ignorant about politics, they will be unable to deliberate fruitfully. But these problems are not decisive against my account. Chapter 5 investigates how the kind of public dialogue I defend offers important resources for rebuilding trust and goodwill in divided societies. As for political ignorance, Chapter 6 demonstrates that it is equally problematic for other political systems, democratic and non-democratic. Since this ignorance threatens all accounts, the solution is not to abandon inclusive deliberation, but rather to tackle political ignorance itself.

TABLE OF CONTENTS

Preface	i
Introduction	1

Part 1. Motivating the Project

<i>Chapter 1. Democracy and Division: Diagnosis of a Problem</i>	13
--	----

Part 2. Norms of Democratic Public Deliberation

<i>Chapter 2. Shared Reasons and the Exclusiveness of Public Discourse</i>	47
<i>Chapter 3. Countering Public Hate Speech</i>	85
<i>Chapter 4. Rage Inside the Deliberative System</i>	117

Part 3. Attitudinal Presuppositions of Public Deliberation

<i>Chapter 5. Deliberation Without Goodwill</i>	159
<i>Chapter 6. Deflecting the Problem of Public Ignorance</i>	187
Conclusion	216

Bibliography	220
---------------------------	-----

PREFACE

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

I have benefited enormously from the insights, advice, and encouragement of mentors, friends, and family while writing this dissertation. Thanks first and foremost to my two supervisors, Clare Chambers and Rae Langton, for each offering an ideal combination of support and critical feedback throughout. I am also particularly grateful to Jess Kaplan, who has read, discussed, and commented brilliantly on each chapter. I could not have hoped for a better political philosopher to exchange with on the doctoral programme.

For invaluable help on written drafts and/or innumerable discussions, thanks to Gabriele Badano, Matthias Brinkmann, Fabien Cante, Karamvir Chadha, Jennifer Daigle, Luke Davies, Fiona Doherty, Matt Dougherty, Tom Dougherty, Cécile Fabre, John Filling, Rachel Fraser, Alexander Greenberg, Steven Gubka, Richard Holton, Samuel Hughes, James Hutton, Matt Kramer, Nakul Krishna, Hélène Landemore, Cathy Mason, Mihaela Mihai, Alfred Moore, Avia Pasternak, Katie Robertson, Paul Sagar, Ian Shapiro, Robert Simpson, Shyane Siriwardena, Jason Stanley, Tamás Szegedi, Jens Van 't Klooster, Ralph Weir, Wes Wrigley, and Yuan Yuan. Thanks also to audiences at the Cambridge Political Philosophy Workshop, Cambridge Graduate Seminar, Cambridge Forum for Legal and Political Philosophy, Cardiff Workshop on

Religion, Hate, and Offence in a Changing World, and Yale Moral Philosophy Working Group.

I am deeply grateful to the Arts and Humanities Research Council for their financial support, which included a full scholarship for doctoral study and a Research Training Support Grant, as well as to Alistair Swiffen for helping me organise a study visit at Yale.

Finally, I couldn't have completed this project without the perceptive criticism, sound advice, and unwavering encouragement of Joanna Demaree-Cotton, and the endless support of my brothers and parents.

INTRODUCTION

1. Democratic Discourse in Divided Settings

Profound and wide-ranging divisions are a defining mark of contemporary politics. People disagree not just about what constitutes a good and fulfilling life, but also about the basic demands of justice, and how these should be satisfied.¹ These ethical divisions run deep. They are rooted, partly, in the fact that different people who belong to different social groups often experience different facets of the social world.² They work different jobs, reside in different neighbourhoods, make different acquaintances, and, accordingly, encounter different and often highly unequal challenges. Some are used to giving orders, others to taking them. Some are respected and admired, others ignored or vilified. Some are real estate magnates, others persistently struggle to pay their rent. With such different and partial experiences of the way the world is, people readily form conflicting views regarding how it ought to be changed.

But ethical divisions do not simply result from contrasting experiences of the observable world. In addition, such divisions are consolidated by people's broader philosophical commitments—for instance, metaphysical commitments regarding what kinds of beings populate the world.³ While a religious believer might hold that acts are made right or wrong by the divine commands recorded in biblical texts, atheists would staunchly disagree. So, even if they had the same sensible experiences of the world, and agreed about what empirical facts obtained, their broader philosophical disagreements might yet lead them to different moral convictions.

Contemporary societies, then, standardly involve wide-ranging and deeply-rooted ethical divisions, which are partly sustained by social inequalities that give rise to contrasting experiences of the world. These inequalities often seem grossly unjust. Poor and uneducated citizens, for instance, regularly risk being exploited by their employers. Similarly, racial or religious minorities are commonly vulnerable to mistreatment or

¹ I follow Raz (1998, 47), Waldron (1999, 105-106), Dryzek (2010, 89-91), Talisse (2009, 3-4), and Enoch (2006, 24) in claiming that contemporary societies involve disagreement not just on the good, but also on the basic demands of justice. This is consistent with Rawls's (2001, 32-38) claim that there can be an "overlapping consensus" on justice, since that claim applies only to reasonable comprehensive doctrines.

² On the epistemic obstacles, including experiential obstacles, that contribute to disagreement, see Young (2000, 115-120) and Rawls (2001, 35-36).

³ Valentini (2013, 181-182).

oppression at the hands of the majority. More generally, disempowered groups often live in a condition of domination: their fates depend on the arbitrary decisions of far more powerful forces. How should we go about realising justice in such non-ideal circumstances? In other words, by what means should we strive to eliminate the unjust social inequalities that plague existing societies, when these societies also involve wide-ranging and deeply-rooted ethical divisions?⁴

Democratic politics, with its emphasis on inclusive public speech, lies at the heart of the answer. In the struggle against injustice, one of the most central practices involves agents bringing forward their concerns, revealing their experiences, denouncing abhorrent conventions, and demanding public justifications of others. Examples abound. Legislators passionately defend same-sex couples' right to marry. Judges publicly condemn redrawing political districts to disenfranchise particular social groups. Whistle-blowers expose the corruption of public officials. And citizens take to the streets to oppose, loudly and resolutely, wanton police violence against racial minorities.

Political philosophers have long acknowledged this insight in their normative theorising. For the past three decades, the dominant tradition of democratic theory—deliberative democracy—has placed inclusive processes of public discourse at the heart of the democratic ideal. While the universal and equal right to vote is of course an essential feature of democratic decision-making, deliberative theories of democracy insist that exchanging reasons in the public sphere is no less central.⁵

And yet, the deliberative ideal of democracy has come under sustained attack from a host of critics who contend that it is too idealised.⁶ According to this challenge, standard formulations of this ideal advance a vision of inclusive public discourse that abstracts away from the pervasive divisions and unjust power asymmetries that characterise actual societies. Consequently, the ideal is too distant to guide our actions in the near future. Perhaps democratic deliberation is a powerful way of realising justice when participants share some ethical common ground, have similar experiences of the world, and like each other. But this tells us very little about the justice-promoting powers of democratic discourse in divided contexts like our own, where public speech is routinely used to degrade others, where speakers rage at other citizens, and where

⁴ For discussion of the conception of justice I am presupposing, see Chapter 1, Section 2.

⁵ For early accounts, see Manin (1987) and Elster (1997). Mansbridge et al (2010) supply an overview.

⁶ E.g., Posner (2003), Shapiro (2003), Green (2010), Somin (2013), Parvin (2015), Brennan (2016), Achen and Bartels (2016).

participants are sometimes neither knowledgeable nor goodwilled enough to learn from the input of others. In a searing indictment which captures this sceptical outlook, Richard Posner concludes that deliberative democracy is “a pipe dream hardly worth the attention of a serious person”.⁷

While many are tempted to jettison the deliberation-centred ideal of democracy, the prominence of inclusive public speech in historical and contemporary struggles for justice suggests that doing so would be a mistake. Instead, what we must do is adapt our philosophical models of inclusive deliberation to make them more relevant. This is the challenge that I will be taking up in this inquiry. What would democratic public discourse have to be like, we should ask, for it to be as close as possible to non-ideal conditions, with their divisions and injustices, while remaining an effective tool for realising justice? To put this question slightly differently: what are the minimal conditions—or ‘threshold’ conditions—that must hold for inclusive deliberation to be a promising way of combatting injustice?

Various types of conditions bear on the effectiveness of public deliberation. Its effectiveness at advancing justice depends, among other things, on what norms regulate (formally or informally) the content and form of public deliberation, what symbolic and material inequalities obtain between participants, and who counts as a participant in the first place. After motivating this project (Part 1), I will primarily be concerned with the issue of what norms should regulate democratic public deliberation (Part 2). However, I will also investigate the symbolic and material conditions that are most intimately bound up with the successful implementation of these discursive norms—in particular, I will consider what attitudes and dispositions these norms require participants to possess (Part 3).

What is a discursive norm? Broadly understood, it is a conventional rule that regulates or constrains speech. To illustrate, consider some of the norm-related questions I will be examining. Must deliberative participants appeal to reasons that are shared by other participants? Do expressions of immoral or degrading views have any place in democratic discourse? And even supposing that a speaker’s public utterance expresses considerations that are morally appropriate and shared, should she nevertheless refrain from voicing those considerations with negative emotions, such as anger or resentment? As these examples suggest, discursive norms can be more or less

⁷ Posner (2003, 163).

formal, depending on how strong the moral reasons underpinning them are. For instance, because degrading speech is commonly taken to engender deeply problematic harms, the salient question is whether it should be restricted by formal legal rules, which impose coercive sanctions on offenders. By contrast, hardly anyone suggests legally restricting the expression of profound anger *per se*, or legally prohibiting appeals to non-shared reasons. Instead, the relevant norms in such debates tend to be informal rules, which are either unenforced or enforced less intrusively, for example through verbal expressions of disapproval.

Although the question of what norms should regulate democratic public deliberation in non-ideal contexts is pressing, and although critics of deliberative theories of democracy have long lamented their irrelevance to such contexts, democratic theorists have yet to address this question in a sufficiently sustained and integrated way. A natural place to look might be the vast literature on public reason. Public reason theorists have developed extremely detailed analyses of the kinds of reasons that are capable of legitimating political power. However, their primary focus in fact tends to differ from my own. In the first place, many public reason theorists simply supply standards for the public *justifiability* of political power, while maintaining that these standards needn't constrain actual deliberative processes of justification.⁸ Furthermore, public reason theorists who *do* supply standards for actual deliberative processes—for instance, Rawlsian public reason theorists—tend to be primarily interested in ideal theory. That is, they investigate what norms should regulate public deliberation in contexts that are substantially better than actual.⁹ To be clear, this is not to say that public reason theory is irrelevant to my inquiry, or that its prescriptions are necessarily inappropriate in actual circumstances. Rather, the point is that it remains an open question—which I will explore—to what extent the detailed prescriptions public reason theorists make for ideal theory are also valuable in non-ideal conditions.

What about philosophical work on public discourse that concerns itself with non-ideal conditions? On the one hand, much of it is driven by specific practical problems. Consequently, it tends not to be integrated with a broader account of democratic discourse. Consider, for example, examinations of whether legal restrictions should

⁸ E.g., Vallier (2015, 155-156), Gaus and Vallier (2009, 65-66), Lister (2008, 281-282).

⁹ E.g., Rawls (1997), Quong (2011, esp. 139, 143; 2017, Section 8), Joshua Cohen (1989), Estlund (2008).

curtail degrading hate speech,¹⁰ and discussions of anger's political role.¹¹ As we will see, these examinations have developed crucial conceptual resources for understanding the moral concerns associated with hate speech and angry speech. However, they are commonly carried out in isolation from broader theories of democratic public deliberation. What is lacking here, then, is a systematic account of how the particular norms being defended should fit together with other norms regulating democratic discourse.

On the other hand, many democratic theorists concerned with non-ideal conditions face the opposite problem: they defend the value of democratic discourse broadly construed in non-ideal circumstances, but without closely examining specific discursive constraints. Most prominently, “agonistic democrats” such as Iris Marion Young emphasise the key role played by inclusive speech in opposing unjust power structures. As such, agonistic theories of democracy are acutely sensitive to the grave power asymmetries and divisions that characterise non-ideal societies. But agonistic democrats seldom explore what particular norms should govern the content of public discourse at great length. This is partly for principled reasons—namely, that they are often hostile to constraints on the content of democratic speech.¹² Though my investigation is in many respects deeply indebted to Young's democratic theory, I am not in principle hostile to such constraints, and therefore wish to examine them more closely.

Still, we should not overstate the point. Some democratic theorists *have* aimed to develop systematic accounts of democratic public deliberation that offer guidance in divided societies.¹³ My disagreements with them are more substantive in nature, and I will discuss them in the course of my inquiry. Even so, it is worth noting that even Sharon Krause, who has developed one of the richest such theories by combining the insights of moral psychology, democratic theory, and feminist thought, nevertheless offers only a relatively brief discussion of how we should address the degrading speech which plagues existing democracies (compared, say, with legal philosophers and philosophers of language who examine hate speech), and of the distinctive place of anger in public deliberation.

¹⁰ E.g., Waldron (2012).

¹¹ E.g., Nussbaum (2016a), Srinivasan (2017).

¹² Young (2000). See also Mouffe (1999), Tully (2005), Schaap (2006), and Dryzek (2010). On Dryzek's nuanced position, see note 67, Chapter 2.

¹³ E.g., Krause (2008), Williams (2000), Gutmann and Thompson (1996), Chambers (2009), Mansbridge et al (2010).

In light of this overview, we can better appreciate the desiderata that my account of democratic public discourse will seek to satisfy. I will articulate and defend a normative account 1) that explores what particular norms should regulate public deliberation, 2) that focuses on non-ideal political contexts, with their divisions and injustices, and 3) that gives a sense of how these particular norms should be integrated with one another. In the process of doing so, I will strive to bring into closer dialogue the valuable strands of research cited above.

I will defend the following account. As a means of preventing domination, participants in formal deliberative arenas should appeal to shared considerations. However, because little is shared between participants in divided societies, this first norm risks excluding important considerations from public speech. To offset these exclusionary tendencies, public deliberation should simultaneously give an important role to emotionally-charged and non-argumentative forms of speech, such as narrative. While this last suggestion might seem overly *inclusive*, by opening the door for deeply angry narratives and eventually for narratives expressing degrading or hateful views, we should not wholly reject this implication. Anger-infused narratives, when used skilfully, perform a crucial epistemic function in divided democracies, by enhancing our understanding of persistent injustices. As for speech expressing degrading or hateful views, there are strong (though defeasible) grounds for thinking that we should allow and authoritatively denounce it, rather than legally prohibit it.

Now, severe distrust and widespread ignorance may limit the justice-promoting powers of public deliberation, thus understood. Put differently, it might be thought that unless citizens trust each other and satisfy a threshold of knowledge, they will be unable to engage fruitfully with one another. Nevertheless, these problems are not decisive against my account. Firstly, the kind of public dialogue I advocate offers key resources for rebuilding trust in divided societies. As for widespread political ignorance, I will demonstrate that it is at least as problematic for alternative political systems. In this light, we should rather understand the inclusive deliberation I advocate as a much-needed—though imperfect—antidote for ignorance.

Before outlining this argument more fully, an important qualification is needed. I am offering a philosophical framework relating to democratic public deliberation. The primary aim of such a framework is to enable a better understanding of the theoretical possibilities for regulating public deliberation, and of the kinds of costs—moral,

epistemic, and other—associated with them. These theoretical insights can and should incline us towards certain discursive norms. But this stops short of allowing us to make highly detailed prescriptions regarding what specific norms we should implement in specific political settings. Such detailed prescriptions depend on various empirical questions, including, for instance, what the political circumstances are, and whether it is possible to institutionalise given norms under these circumstances. Accordingly, the task of making precise policy prescriptions cannot be performed solely by political philosophers. While philosophy’s conceptual insights contribute to generating theoretical hypotheses that inform and guide empirical research, the specialised knowledge of social scientists and policy-makers is needed to test and refine these hypotheses.¹⁴ In what follows, I will therefore strive to indicate where my argument depends on empirical assumptions, where further empirical research is needed, and in what respects this renders my argument defeasible.

2. Outline

2.1. Motivating the Project

Chapter 1 diagnoses a problem for democracy. It introduces a dilemma that plagues attempts at advancing justice in non-ideal circumstances. Although democratic decision-making should be understood as an attempt to navigate between the horns of this dilemma, I will show that a version of the same dilemma ultimately resurfaces *within* democratic theory, albeit in a slightly less severe form.

In laying out this problem, this chapter serves two main functions. The first is motivational. While democratic decision-making is an improvement on the alternatives, the fact that it does not fully avoid the dilemma it sets out to resolve remains deeply problematic. Now, the more a normative model of democracy requires that pre-existing thresholds of justice be met, the more it runs afoul of this dilemma. This insight gives us powerful reasons to reorient our theorising about democracy and democratic deliberation as follows: we should strive to articulate a conception of deliberative democracy that is as close to non-ideal conditions as possible, while remaining capable of realising justice.

¹⁴ On this methodology, see Estlund (2008, 12-15) and Waldron (2012, 11-17).

The second function has to do with conceptual framing. Framing democratic theory as an attempt to address this initial dilemma 1) will clarify how democratic deliberation is in principle meant to advance justice. At the same time, it 2) will enhance our understanding of how theories of democracy might go wrong, thereby yielding a useful taxonomy of existing deliberation-oriented theories of democracy. Relatedly, keeping the horns of this dilemma in sight will serve to remind us of the pitfalls that must be avoided in subsequent chapters.

2.2. Norms of Democratic Public Deliberation

Chapters 2 to 4 take up the question of what norms, formal or informal, should regulate the content and form of democratic public speech in non-ideal conditions. Chapter 2 considers a commonly-suggested norm, the Shared Reasons Constraint, according to which deliberators must appeal to considerations that are suitably shared. On the one hand, considerations of domination strongly support such a constraint: requiring that deliberation be framed in terms of shared reasons facilitates the contestation of political power, and thereby prevents such power from being exercised arbitrarily. Nevertheless, recent work in social epistemology reveals that this constraint risks being overly exclusive. In divided democracies, different social groups have radically different experiences of the social context. The significance of these experiential divides is amplified by what Miranda Fricker calls “hermeneutical injustices”, injustices that strip individuals of the conceptual resources needed to articulate their meaningful experiences and concerns.¹⁵ In such contexts, even the weakest formulation of the Shared Reasons Constraint excludes some utterances that should have a place in public debate. After examining the limitations of traditional strategies for mitigating this problem, I advance an alternative strategy. To offset the Shared Reasons Constraint’s overexclusiveness, we should give a greater role to non-argumentative and to emotionally-charged forms of speech in public discourse. Emotionally-charged narrative, I suggest, helps to communicate experiences between divided groups. In so doing, it enriches the pool of shared resources from which the Shared Reasons Constraint proceeds, and thereby mitigates its exclusionary implications.

¹⁵ Fricker (2007).

Chapter 3 considers an opposite concern: that combining the Shared Reasons Constraint with discursive norms that invite emotionally-charged personal narratives may be too inclusive. In particular, it may open the door for public assertions of deeply disrespectful views. Focusing on public hate speech, I investigate whether this ‘overinclusiveness’ requires supplementing the norms advanced in Chapter 2 with a coercive legal norm suppressing degrading contributions. I argue against doing so. Defenders of legal bans often operate with an unrefined understanding of what it means to counter hate speech with more speech. Specifically, they overlook the role that the state can play in endowing ‘counterspeech’ with authority. Robust state-sponsored counterspeech, I suggest, constitutes a morally preferable means of defusing the harm in hate speech. Admittedly, in non-ideal conditions, state-sponsored counterspeech remains an imperfect solution to hate speech. However, I will show that the weightiest objections to such counterspeech can be turned, with even greater force, against coercive legal bans. Note, though, that this argument remains defeasible. I focus on Waldron’s account of what makes hate speech harmful, which itself builds on the work of feminist and critical race theorists. While highly influential, this account is not the only one. A fuller case against coercively eliminating expressions of degrading contents from public deliberation would have to demonstrate that the argumentative strategy I deploy can be extended to other kinds of harms. For many of the harms discussed in the literature, this undertaking will be in significant part empirical.

Chapter 4 complements Chapter 2 in another way. While Chapter 2 recommended giving a greater role to emotionally-charged narratives in public discourse, Chapter 4 substantiates this claim by examining angry narratives, specifically. It is commonly held that publicly expressing anger is counterproductive for justice. I resist this challenge by articulating a crucial sense in which angry narrative can be epistemically productive. Drawing on recent developments in the philosophy of emotion, which emphasise the distinctive phenomenology of emotional experience, I contend that conveying anger to one’s listeners can play an indispensable role in alerting them to previously-overlooked injustices, and in enhancing their understanding of these injustices. Finally, I defuse the most powerful objections to this defence, partly by exposing how they overlook the systemic character of democratic deliberation.

2.3. The Attitudinal Presuppositions of Public Deliberation

The previous chapters defend a set of discursive norms that help advance justice in non-ideal conditions. But what symbolic and material conditions must already obtain for these norms to be successfully implemented? In particular, how far does their implementation require participants to have attitudinal dispositions that are actually lacking in the real world, where status-based and material inequalities are rife? The final two chapters consider two attitudinal dispositions: the disposition to extend goodwill and trust to one's interlocutors, and the disposition to seek out and accept reliable information.

Deliberative theories of democracy, some critics have forcefully objected, require more trust and goodwill than obtains in divided societies. Chapter 5 develops a two-pronged response to this worry. First, drawing on the systemic approach to deliberation, I argue that the effectiveness of public deliberation at advancing justice demands less goodwill than the objection suggests. Second, even if it were too demanding, features of public deliberation that are widespread in non-ideal conditions—namely hypocrisy, anger, and even the occurrence of abhorrent speech—have properties that can be harnessed to rebuild goodwill in divided societies. This might seem counterintuitive: *prima facie*, these features of public speech seem wholly at odds with trust and goodwill. But I will contend that this commonly-held intuition should be challenged. Importantly, moreover, the properties of non-ideal democratic discourse that help generate goodwill are consonant with the discursive norms that, on independent grounds, I defended as means of realising justice. Hence, this defensive exploration lends further support to my normative picture of democratic public deliberation.

Finally, Chapter 6 addresses the worry that ordinary citizens tend not to be disposed to seek out and accept reliable information. The resulting political ignorance constitutes one of the most daunting obstacles to using democratic public deliberation to combat injustice. This is because the justice-promoting powers of deliberation typically depend on participants already satisfying some threshold of knowledge. Nonetheless, I argue that public ignorance is not a special problem for deliberative theories of democracy. On the contrary, ignorance proves equally problematic for the normative political models, democratic and non-democratic, which are touted as remedies. Thus, departing

from deliberative democracy on this basis is misguided. Far from being the source of the problem, inclusive deliberation is best understood as an attempted solution to the difficulties stemming from widespread ignorance. Given that these difficulties are ubiquitous, and given the promise of deliberative innovations, both conceptual and institutional, we must try and make this solution work.

PART 1

MOTIVATING THE PROJECT

CHAPTER 1

DEMOCRACY AND DIVISION: DIAGNOSIS OF A PROBLEM

1. Introduction

What is justice? This question has preoccupied a wide range of political philosophers, historical and contemporary. Just as Plato devotes most of his *Republic* to identifying the essence of justice, in the city and in the soul, so Rousseau's aim in *The Social Contract* is to delineate a political order that reconciles justice and utility.¹ Similarly, Rawls asserts at the outset of *A Theory of Justice* that, since "justice is the first virtue of social institutions", he intends to develop a systematic picture of justice in an idealised society.²

However, a second question lurks in the background of these philosophical projects: given what justice is, how should we go about realising it, from where we are? Take Plato. Having depicted the just city, he considers how it might actually be brought about. Plato suggests that justice cannot be achieved unless philosophers rule.³ Unfortunately, as he concedes, the corrupt nature of actual politics makes it nearly impossible for philosophers to access power without compromising their motivation to realise justice.⁴ Rousseau likewise acknowledges the problem without offering a satisfying solution. Achieving his ideally-just state requires instituting an egalitarian order. But since hierarchical institutions have warped the moral psychology of actual people, they are unlikely to be able, let alone willing, to realise such an order. Rousseau ultimately invokes the ominous figure of a godlike legislator, who supplies imperfect beings with laws constituting the just political order.⁵

These two cases suggest that, without necessarily being their primary focus, the question of how to promote justice in the actual world has long puzzled theorists of justice. In recent years, this puzzle has not only persisted, but has become a central concern of political philosophers. Charles Mills, for example, has influentially argued

¹ Rousseau (1973[1762], 181).

² Rawls (1999, 3).

³ Plato (2007[c. 380 BC], 473c-e).

⁴ Ibid., 490e-496e.

⁵ Rousseau (1973[1762], 194-197).

that by focusing on ideal justice in his *Theory*, Rawls ends up neglecting the question of how justice should be advanced in our world. Unless political philosophy gives greater emphasis to the latter question, Mills continues, it risks being irrelevant to our most pressing political concerns, and eventually constituting a counterproductive distraction.⁶

A major driving force underlying this trend is the growing emphasis by political philosophers on the profound divisions that characterise contemporary societies.⁷ As we saw earlier,⁸ contemporary non-ideal societies are divided over very many ethical matters. Individuals disagree not only regarding what the good life involves, but also regarding the basic demands of justice. These divisions are reinforced, in part, by pervasive social inequalities. Different groups occupy different and often highly unequal social positions, and experience different and often highly unequal constraints. Consequently, they frequently come to view the world in radically different ways, and form contrasting opinions regarding how we should change it.

These deep divisions have made the question of how to go about realising justice more salient, partly because they give rise to a more philosophically challenging version of that question:

The Promotion Problem: How should we go about promoting justice, given that actual non-ideal conditions are characterised, among other things, by wide-ranging disagreement on what justice involves?

How is this version more challenging? If everyone agreed on what justice means, the question would reduce largely to the problem of identifying the most efficient means of advancing justice. Many (though not all) challenges involved in doing so concern social scientists rather than political philosophers. By contrast, starting from a situation of ethical division raises difficult philosophical issues, including: what are the moral costs of exercising power over others in ways that they find unacceptable? Who should exercise that power? And what is the appropriate epistemic attitude to persisting moral disagreement?

⁶ Mills (2005, 168). For how the question of advancing justice has taken centre-stage within the ideal/non-ideal theory debate, see also Sen (2009) and Gaus (2016).

⁷ For examples of this trend, see Raz (1998), Waldron (1999), and Valentini (2013).

⁸ See Introduction, including notes 1-3.

The present chapter examines the highly influential view that, in such non-ideal and divided settings, the appropriate way to go about promoting justice is via democratic decision-making. While I strongly sympathise with this view—which I call the Democratic Solution—the primary purpose of this chapter is to diagnose a serious problem it encounters. We can put the problem as a dilemma: in non-ideal, divided political settings, *either* existing accounts of democratic-decision-making seem unpromising as ways of promoting justice; *or* they presuppose prior non-democratic exercises of power that are morally and epistemically suspect, as well as insensitive to background contexts of power.

Importantly, the purpose of this argument is not to dismiss democracy as a way of promoting justice. It remains, I believe, the best solution we have to the Promotion Problem. Instead, the purpose is to motivate a project (which the rest of the thesis pursues) of reconceptualising the democratic ideal for non-ideal conditions, so as to minimise the extent to which it runs afoul of the above difficulties.

The argument will proceed as follows. I begin by outlining and critiquing two opposite responses to the Promotion Problem. The first, which recommends simply forcibly imposing justice in divided contexts, faces important moral, epistemic, and power-related difficulties. The second, which responds to these difficulties by adopting a more modest political aim than justice (such as peace or order), is too pessimistic. These two approaches constitute the two unsatisfying extremes between which the Democratic Solution aims to navigate (Section 2). With this initial dilemma in place, I introduce the Democratic Solution. The inclusive deliberation it prescribes, I suggest, seems a *prima facie* promising way of realising justice. Nevertheless, I then argue that the Democratic Solution ultimately faces a serious problem: a version of the dilemma we started with resurfaces *within* democratic theory. Indeed, either democratic decision-making presupposes non-democratic uses of force that generate moral, epistemic, and power-related difficulties, or it seems ill-adapted to promoting an ambitious aim such as justice (Section 3). I conclude by examining what this means for the democratic promotion of justice in non-ideal settings. While this chapter aims to dampen some of the optimism pervading democratic theory, it is not wholly negative. I offer this diagnosis of democratic theory's ills not to pronounce it dead, but to motivate the search for an adequate cure (Section 4).

2. Motivating the Democratic Solution: A Dilemma

The present section raises and critiques two initial responses to the Promotion Problem. Before proceeding, an important clarification is needed. Some of the arguments I will offer when critiquing these responses appeal to intuitions regarding what justice involves. For instance, I sometimes presuppose that oppressive hierarchies and domination are *prima facie* unjust. But if I am myself relying on a contentious understanding of justice, how can I hope to convince those who do not share these views? Aren't I ignoring the pervasive disagreements about justice discussed in the introduction?

To address this worry, let me clarify my aim. My intention is not to convince absolutely everyone that we should go about promoting justice in a certain way despite vast disagreements about the meaning of justice. Instead, insofar as my arguments rely on disputed intuitions about justice, I intend to show something more modest: that *if* you share these moral concerns, you are committed to certain ways of promoting justice in divided contexts.⁹

More specifically, in this and subsequent chapters, I will be relying on the following broad egalitarian intuitions about justice. Justice, I am assuming, requires treating human beings with respect as free and equal persons. Positively, this means upholding human beings' basic capacity to lead autonomous lives, whereby they consider, commit to, and pursue goals that are dear to them. Negatively, this requires eliminating hierarchical social relations that prevent human beings from meaningfully exercising that capacity. Thus, I am assuming that unequal social relations such as domination (roughly: being at the mercy of others' arbitrary decisions) and oppression (roughly: being denied basic opportunities to pursue one's cherished goals or to develop one's skills) are *prima facie* unjust.

Over the course of my argument, I will elaborate on specific features of this picture of justice, particularly the idea of domination. In the meantime, it is worth noting that although these intuitions about justice are inevitably contested by some citizens of contemporary democracies, they are consistent with a wide range of philosophical

⁹ Talisse (2009) attempts to achieve this more ambitious aim by appealing only to non-ethical epistemic premises. Similarly, Waldron (1999, 160-161, 298-301) strives to avoid relying on disputed ethical premises by appealing to pragmatic considerations. For persuasive criticisms, see Ahlstrom-Vij (2011) and Fabre (2000b, 276-282).

views about justice, both inside and outside of the liberal tradition.¹⁰ Having clarified this, let us now consider two opposing answers to the Promotion Problem.

2.1. The Moral Imposition Solution

An initial answer to the Promotion Problem runs as follows:

The Moral Imposition Solution: In conditions of wide-ranging ethical disagreement, insofar as one has the power forcibly to impose one's views regarding what justice demands, one should simply exercise that power.

By 'simply', I mean that the use of force is not constrained by prior collective procedures, such as the characteristically democratic procedures of having all those concerned vote on the issue, or of offering justifications for one's use of force to those subjected to it and allowing them to respond.

This solution calls to mind Plato's philosopher king and Rousseau's legislator. But a better example in our context is Augustine. Augustine explicitly asks how we should react to moral disagreement between conflicting religions. In response, he affirms the need forcibly to compel heretics to comply with justice, as defined by what he deems the true religion.¹¹ "Why shouldn't the church [...] force her lost sons to return", Augustine asks, "when her lost sons have themselves compelled others to be lost?"¹² As the rhetorical question emphasises, there is something naturally appealing about this approach: if justice matters greatly, and if one has the power to do so, why not simply forcibly impose justice on dissenters?

A more recent example, which might seem less alien to modern ears, is Young's activist. In "Activist Challenges to Deliberative Democracy", Young stages a dialogue

¹⁰ In the liberal tradition, these intuitions are closest in spirit to Elizabeth Anderson's (1999, 313) "relational" egalitarianism, which "seek[s] a social order in which persons stand in relations of equality [...] as opposed to a hierarchical one".¹⁰ But they are also consistent with more "distributive" liberal egalitarian views of justice. According to these views, justice is not essentially about equalising social relations. Rather, it is about equalising distributions of opportunities for welfare, resources, or advantage. E.g., Arneson (1989), Dworkin (1981b), G.A. Cohen (1989). Despite this disagreement with Anderson, however, distributive views typically acknowledge that opposing hierarchical relations (such as domination or oppression) may be *instrumentally* necessary for justice. E.g., Arneson (2000, 341-342). Finally, this picture of justice is consonant with the views of some theorists who are more sceptical of the liberal tradition. Young (1990, 37-38), for example, explicitly affirms that justice requires eliminating relations of domination and oppression.

¹¹ Augustine (2001[c. 147], 178).

¹² Ibid., 187-188.

between an activist and a deliberative democrat. Recall that deliberative democracy (which Section 3 will outline more extensively) is a conception of democracy that gives a central role to inclusive public dialogue. Young's activist rejects the idea of engaging in such dialogue, which he deems naïve in contexts involving entrenched divisions. Instead, he recommends promoting justice through direct action. Direct action involves disruptive behaviour (including sit-ins or demonstrations) which might result in breakage, and perhaps even violence (in self-defence).¹³ Young's activist illustrates the Moral Imposition Solution in two ways. First, he responds to disagreement by engaging in the direct, forcible promotion of what he believes is just. Second, in doing so, he eschews democratic procedures of exchanging justifications concerning his use of force.

Now, Young acknowledges that her activist represents a valuable perspective, but nevertheless an incomplete one. Her aim is less to defend this activist wholeheartedly, and more to make democratic theorists engage more fully with his pragmatic concerns.¹⁴ More generally, while some contemporary theorists are closer to the Moral Imposition Solution than others—in particular, those who ascribe little importance to democratic procedures in legitimating power¹⁵—the 'pure' Moral Imposition Solution is largely rejected by contemporary theorists.

But although this approach to promoting justice is unpopular among most contemporary philosophers, engaging with it is fruitful given my broader aim. My aim, as advertised in the introduction, will be to show that the strategy of advancing justice via democratic processes—the Democratic Solution—ultimately cannot fully escape the difficulties which afflict the Moral Imposition Solution. In pursuing this aim, engaging with a stylised opponent is helpful. First, it helps bring out clearly the difficulties that the Democratic Solution will be introduced to avoid. Moreover, it highlights just how problematic it is that it inescapably runs afoul of some of them: the fact that, as I will argue, the Democratic Solution is vulnerable to similar problems as such an unpopular view should worry democratic theorists. With this perspective in mind, let us consider three problems with the Moral Imposition Solution.

¹³ Young (2001, 671-674).

¹⁴ *Ibid.*, 688.

¹⁵ E.g., Brennan (2016), López-Guerra (2014, ch. 2).

a) *The Domination Problem*

The first objection holds that simply forcibly imposing one's conception of justice on others dominates them. And since, I am assuming, domination is unjust, this means that the current strategy for advancing justice is itself *pro tanto* unjust.

Domination, Philip Pettit explains, consists in subjection to "alien control".¹⁶ B is subjected to A's alien control when A has "the capacity, not subject to [B's] direct or indirect check, to interfere in [B's] choice".¹⁷ That is, what options B has access to depends on A's arbitrary or unchecked decisions. Consider two notable features of domination. First, having many good options available is insufficient for non-domination. The slave of a benevolent master is dominated because, although he has many good options, the master could arbitrarily interfere with his options if she wanted. Second, having many options is not necessary for non-domination. This is because one can be subjected to power which restricts one's options but which is not alien. This is the case of power which, because it is subject to checks, is forced to track one's concerns.¹⁸

Returning to moral imposition, when A simply unilaterally forces her view of justice on B and is sufficiently powerful to do so, A is not subjected to systematic checks on B's behalf. Consequently, A's exercise of power is not forced to track B's ethical concerns. Thus, B is subjected to A's alien control. This domination is potentially quite serious. Insofar as B's disagreement with A's view of justice results from her deep-seated ethical commitments, A is exercising alien power over aspects of B's life that are central to her personal integrity.

Now, such domination *could* conceivably promote non-domination overall. Indeed, unchecked power might be used to establish a political order involving more accountable and just institutions. My point is that there are nevertheless strong *pro tanto* reasons based in the value of non-domination to seek an alternative to the Moral Imposition Solution: even if forcible imposition sometimes promotes non-domination (and therefore justice) overall, it would be preferable to find a way of promoting justice that is not itself dominating. Put differently, if a just state of affairs could be realised through a relatively peaceful strategy rather than through unchecked force, that would

¹⁶ Pettit (2008, 102-103).

¹⁷ Lovett and Pettit (2009, 14).

¹⁸ Pettit (2008, 116-118).

be morally preferable. Perhaps it will turn out that no such strategy exists. However, as Section 3 will discuss, democratic theorists think one does.

b) The Epistemic Problem

The Moral Imposition Solution also seems inadequately sensitive to the epistemic obstacles individuals encounter in divided societies. Laws, social norms, and infrastructures often subject different social groups to different constraints. Accordingly, different aspects of society are visible to different groups.¹⁹ Sexual harassment in the workplace is likely to be comparatively more visible to women. Likewise, working-class inhabitants of Paris's impoverished suburbs are likely to know more about housing conditions there than inhabitants of its affluent suburbs. Because of this asymmetric distribution of information, it is difficult for any single agent, who necessarily occupies a particular social position, to gather all the information relevant to fully understanding the social context.

In the first instance, these epistemic obstacles concern *non-moral* facts about the social context. But non-moral facts bear on *moral* judgments about justice in at least two ways. First, consider that such judgments come at all levels of generality. Some are general principles ('justice demands that each have the opportunity to live autonomously'), whilst others are more specific ('justice demands building affordable housing'). Specific judgments are often derived by combining a general principle with a non-moral fact ('without affordable housing, individuals cannot live autonomously').

Second, attention to non-moral facts can also influence one's general principles. Moral principles, even highly general ones, almost invariably rely on factual presuppositions. A principle demanding that everyone have the opportunity to live autonomously presupposes that people can in principle live autonomously. More relevant to our purposes, imagine someone who never experienced poverty and its social determinants, let alone workplace harassment and discrimination. Having experienced a tight correlation between individual effort and reward, he may be more likely to arrive at the general view that considerations of desert and personal responsibility are central to distributive justice.

¹⁹ Anderson (2006, 11); Young (2000, 112-120).

Thus, what non-moral facts one has ready access to contributes to shaping one's ethical perspective. In turn, what non-moral facts are readily available depends significantly on one's social position. So, ethical divisions at least partly reflect the epistemic partiality of different social positions. This conclusion, Robert Talisse notes, recommends a certain epistemic modesty in the face of ethical disagreement.²⁰ We should take seriously the possibility that people disagree with us not because they are incompetent, but because they have relevant experiences that we lack. The Moral Imposition Solution sits in tension with such epistemic modesty: it prescribes responding to disagreement by unilaterally and forcibly imposing one's views about justice.

Admittedly, epistemic modesty may sometimes be inappropriate. Some social groups may be epistemically privileged concerning what justice requires in *particular* areas of social life. While inhabitants of poor Parisian suburbs may have incomplete perspectives on society as a whole, they may be objectively better, *ceteris paribus*, at knowing how the housing crisis in these parts of Paris should be tackled. In such cases, it may be unwarranted for them to exercise epistemic modesty in the face of disagreement. To handle such cases, the final difficulty the Moral Imposition Solution encounters will apply regardless of whether one strongly believes, with good reason, that one is right about what justice demands.

c) The Context of Power Problem

The final challenge starts from the observation that the Moral Imposition Solution to the Promotion Problem is question-begging. Suppose I am a party in a moral dispute. The Moral Imposition Solution, as a practical guide to action, tells me that I should forcibly impose justice. Doing so requires me to endorse an answer to the question of what justice demands. But in the divided contexts under consideration, what justice demands is precisely what is in dispute.

Why is this problematic? Because the Moral Imposition Solution does not resolve the dispute at hand, but instead involves presupposing what is in dispute, it justifies forcible imposition not just by me, but by my opponent too. When my opponent turns

²⁰ Talisse (2005, 428-430).

to this practical rule, it tells *her* to impose what *she* deems just. Thus, the Moral Imposition Solution can be equally deployed by all parties to the dispute.²¹

This conclusion might leave some relatively unfazed. Such indifference, I suspect, results from the following tendency: political philosophers reasoning about moral issues often place themselves in the position of the state deciding what to do. Once one does this, the problem might seem unimportant. If I am right, and have state power on my side, it matters little whether others want to forcibly impose their views.²² Only I will actually be able to do so.

But things look different when considering the non-ideal contexts in which we live. In practice, who is actually right and who has power often come apart. Across Europe and the US, for instance, many—sometimes majorities—have come to assent to far-right conceptions of immigration justice, conceptions which sometimes recommend the oppressive deportation or mistreatment of migrants. Relatedly, Young concedes that the power her activist exercises in forcibly promoting justice is “a mere David to the Goliath of power wielded by the state and corporate actors [...] he opposes”.²³ In conditions such as these, the fact that the Moral Imposition Solution can be deployed by all is worrying. If those who are right about justice are not the most powerful, and if the Moral Imposition Solution legitimates imposition by all disagreeing parties, then its expected result will not be justice. It will be whatever the powerful take justice to mean, however misguidedly. Thus, the Moral Imposition Solution seems dangerously insensitive to non-ideal contexts of power.

Perhaps this critique reveals that I have misconstrued the Moral Imposition Solution. Young’s activist, for instance, might respond that not everyone is permitted to forcibly impose their conception of justice. Rather, he, unlike others, may do so because he is promoting what justice *actually* requires.²⁴ This reply suggests adjusting the Moral Imposition Solution as follows:

*The Moral Imposition Solution**: In conditions of wide-ranging ethical disagreement, insofar as one has the power forcibly to impose one’s views regarding what

²¹ I am grateful to Chandran Kukathas for pressing this objection. Waldron (2006, 1371) articulates a related objection.

²² E.g., Kolodny (2014, 203).

²³ Young (2001, 675)

²⁴ Ibid.

justice demands, one should simply exercise that power, *provided one holds correct views about justice*.

This version, the response concludes, is not ambiguous: it recommends imposition *only* by those who are right.

Nevertheless, the reformulated version remains just as ambiguous for all practical purposes. Since each party believes that they are actually right about justice, each is bound to infer from the principle that they should impose their conception of justice. What this suggests is that in context, any party—even the correct one—that claims that the principle legitimates their interference but not that of others is begging the question, by assuming what was originally in dispute: that they are actually right about what justice means.²⁵ In practical context of disagreements, then, the principle can still be deployed by all parties. Hence, this reformulation fails to rescue the Moral Imposition Solution.

At this point, one might suggest another amendment to the Moral Imposition Solution. On this view, it should be understood not as what G.A. Cohen calls a “rule of regulation”, but rather as a “fundamental principle”. The former, for Cohen, are practical rules we adopt to guide action in social contexts, with a view to achieving just results. By contrast, a fundamental principle is a standard that tells us about the basic nature of justice. So, fundamental principles allow us to discern whether a state of affairs is good or bad from the standpoint of justice.²⁶ Thus interpreted, the Moral Imposition Solution* asserts that it would be best, from the standpoint of justice, if only those who are actually right imposed their views. But it does *not* say that explicitly applying this principle in divided contexts will help realise this just outcome. In light of this reinterpretation, the challenge under examination seems misplaced.

This reply, however, saves the Moral Imposition Solution only at the cost of rendering it unsuited to our purposes. As explained in the introduction, the Promotion Problem does not ask what the basic nature of justice is. Rather, it asks what standards should be collectively followed to advance justice in divided settings. Accordingly, in

²⁵ Imagine what A, who is actually right, would say to her opponent B. A: ‘By the Moral Imposition Solution*, only I may impose my conception of justice.’ B: ‘Why?’ A: ‘Because I am actually right about what justice requires.’ B: ‘That’s where we disagree. I still think that *I* am right. So unless you give me a new reason for thinking otherwise, you haven’t given me a reason to think that the Moral Imposition Solution* legitimates your imposition rather than mine.’

²⁶ Cohen (2008, 263-272).

Cohen's terminology, it seeks to identify a rule of regulation that offers guidance in non-ideal conditions.

2.2. The Realistic Solution

The Moral Imposition Solution, then, encounters three significant problems. Because these seem daunting, some philosophers adopt an opposite response to the Promotion Problem:

The Realistic Solution: In conditions of wide-ranging ethical disagreement, one should refrain from trying to achieve a highly contested ideal (such as justice), and instead strive to achieve a more modest and uncontroversial ethical ideal (such as peace).²⁷

In what sense is this a *realistic* solution? According to its proponents, peace, by contrast with conceptions of justice or of the good life, is a thin enough ethical aim to be universally endorsed.²⁸ Put differently, peace is a lowest-common-denominator ethical ideal.²⁹ Because this consensus exists, peace can be pursued without encountering the three challenges developed above. Hence, this solution is 'realistic' in the following sense: in response to the realities of intractable disagreement and power imbalances, it adopts an aim that is more readily and unobjectionably realisable than justice.

To evaluate this solution, let us examine the concept of 'peace' more closely. On what Fabian Wendt calls the "ordinary" understanding of peace, peace is a compromise between conflicting parties that aims to establish stable, non-violent coexistence between them.³⁰ The main problem with this interpretation of the Realistic Solution is that giving up on justice to pursue ordinary peace instead has highly unpalatable implications. While all parties to the compromise share an interest in achieving non-

²⁷ E.g., Kukathas (2003, 252), Gray (2000, 5-6), Horton (2010, 439), Wendt (2013, 576).

²⁸ Wendt (2013, 580).

²⁹ The Realistic Solution is not identical to what is sometimes called "political realism". While some political realists—for instance, John Horton—might endorse the Realistic Solution, Wendt (2016) notes that the two can also come apart. To see why, consider that, according to Rossi and Sleat (2014), one of political realism's central tenets is that politics is an autonomous sphere, whose distinctive norms are (at least partly) independent of moral norms like justice. By contrast, the Realistic Solution in principle encompasses the view that the aim of politics is to promote *lowest-common-denominator justice*. Indeed, its only essential feature is refusal to impose highly contested ethical ideals.

³⁰ Wendt (2013, 577).

violent coexistence, the balance of power between them is often very unequal. Therefore, stronger parties can extract greater concessions from weaker parties when agreeing on the terms of peace. Weaker groups might have to abandon key ethical practices or retain them only at the cost of living in a hierarchical order where stronger parties dominate them.

To illustrate, consider the 1598 Edict of Nantes, which marked the end of the French wars of religion, and whereby French Catholics adopted a more relaxed position towards Protestants. While it allowed Protestants to practise their religion, it also upheld Catholicism's ascendancy: Protestantism was geographically restricted, Protestants could not marry, and they were required to pay taxes to the Catholic Church. Against a background of significant inequality, then, ordinary peace risks reproducing pre-existing disparities by institutionalising hierarchical relations. In this light, the Realistic Solution offers too pessimistic and unappealing a conception of politics.³¹ Politics should be more ambitious in its aim.

Proponents of the Realistic Solution might respond by endorsing a more ambitious conception of peace.³² Notably, Chandran Kukathas develops a conception of peace whereby different ethical communities refrain from interfering with each other's internal ethical practices, and leave each other free to live as their conscience dictates. This conception is more ambitious than ordinary peace, as it rejects situations involving institutionalised hierarchical and invasive relations between stronger and weaker groups. Ambitious peace demands radical non-interference between groups.³³

Some have objected that Kukathas's ambitious peace still yields an unappealing picture of politics.³⁴ But in our context, the principal problem is that this understanding of peace constitutes a very disputed ideal. Kukathas's defence of ambitious peace hinges on the claim that individuals have a fundamental interest in living according to conscience.³⁵ Importantly, Kukathas asserts not simply that living according to conscience is valuable. His ambitious peace presupposes that it is *so* valuable that it outweighs reasons to interfere with other groups' internal practices, even when these are oppressive.

³¹ Ibid., 580.

³² Ibid., 581-584.

³³ Kukathas (2003, 126-147).

³⁴ Okin (2002, 214-216).

³⁵ Kukathas (2003, 48-50).

Consequently, Kukathas's ambitious peace runs afoul of the difficulty the Realistic Solution was introduced to avoid.³⁶ The motivation for giving up on justice and pursuing peace instead was that peace is a more modest and uncontroversial ethical ideal. Hence, using political power to pursue it allegedly avoids the difficulties facing the Moral Imposition Solution. But since the robust conception of peace Kukathas offers is a highly contested ideal, recommending that the state enforce this ideal replicates the Moral Imposition Solution's problems.

Generalising from this failure, a difficulty confronts the Realistic Solution. If peace is defined in a truly thin manner, the resulting view of politics is deeply unappealing, and risks conducing to the institutional reproduction of unjust inequalities. If, however, one endorses a more robust conception of peace, it becomes a controversial normative ideal, so that imposing it replicates the problems outlined in 2.1. Admittedly, I cannot rule out in principle the possibility of a conception of peace that is both appealing and uncontroversial. But the task of developing such a conception seems extremely challenging: in deeply divided contexts, all but the thinnest views of peace seem bound to be contested.

2.3. Conclusion: An Initial Dilemma

Sections 2.1-2.2 suggest a general dilemma for promoting justice in non-ideal divided settings. On the one hand, the Moral Imposition Solution seems undesirable in light of three problems: it dominates disagreeing individuals, it is often epistemically inappropriate, and it is question-begging in a way that renders it dangerously insensitive to non-ideal contexts of power. On the other hand, insofar as it is distinguishable from the Moral Imposition Solution, the Realistic Solution yields a highly pessimistic picture of politics. By bracketing justice and embracing a more modest ideal (such as peace) instead, it risks reproducing rather than remediating existing injustices. This may seem an unacceptably high price to pay for avoiding the Moral Imposition Solution's problems.

In what follows, I introduce the Democratic Solution and suggest that, because of democratic public deliberation's instrumental and intrinsic relation to justice, it seems a promising way of navigating between the horns of this initial dilemma. But despite this

³⁶ Wendt (2013, 584-585).

prima facie appeal, I then diagnose a problem for this approach: a dilemma similar to the one that motivates a turn to democracy ultimately resurfaces *within* the Democratic Solution (for a visual representation of this argument, see the diagram at the end of Section 3).

3. The Democratic Dilemma

3.1. The Democratic Solution

How should we bring about justice in non-ideal circumstances characterised by wide-ranging ethical divisions? The most popular proposal today is the following:

The Democratic Solution: In conditions of wide-ranging ethical disagreement, we should promote justice via democratic decision-making processes.³⁷

A collective decision-making procedure is democratic, according to Thomas Christiano, just in case each member of the demos has “an equal say at a crucial stage of the decision-making”.³⁸ Alongside most proponents of the Democratic Solution, I will focus on deliberation-centred conceptions of democracy. Broadly put, ‘deliberative democracy’ refers to the family of democratic theories that place inclusive processes of public reason-giving at the heart of the collective decision-making process.³⁹ On this interpretation of democracy, having an equal say means not just having equal voting power, but also having an equal opportunity to participate, directly or indirectly, in public discussion aimed at influencing practical decision-making.

Before unpacking the notion of democratic public deliberation, note that my aim here is not to exhaustively characterise the best version of deliberative democracy. Prior to doing so, we must clearly understand the problems democratic theory encounters, and this chapter aims to diagnose such a problem. Instead, I now outline characteristics that almost any broadly deliberative conception of democracy accepts, while bracketing

³⁷ See, e.g., Benhabib (1996), Gutmann and Thompson (1996), Mouffe (1999), Waldron (1999), Young (2000, ch. 1), Williams (2000), Urbinati (2000), Bohman (2003), Krause (2008), Talisse (2009), Dryzek (2010, ch. 5), Anderson (2010, ch. 5), Mansbridge et al (2010), Landemore (2013), Shapiro (2016).

³⁸ Christiano (2006, 83).

³⁹ See note 37, this chapter. This broad understanding allows me to include Mouffe (1999) and Young (2000), who sometimes reject the label ‘deliberative democracy’ in favour of ‘agonistic democracy’ (discussed in Section 3.3), under this heading. I return to and criticise non-deliberative conceptions of democracy—e.g., Shapiro (2016)—in Chapter 6.

further controversies—to which I return in subsequent chapters—regarding what precise rules should govern the content and form of deliberation.

Concisely put, public deliberation is the public process whereby individuals mutually offer considerations for and against practical proposals with the aim of determining which proposal is supported by the best reasons. Several features of this definition are worth emphasising. First, the public exchange of justifications, or reasons, is central to deliberative democracy. Moreover, this exchange aspires to transform the opinions of participants so that they align with the best-supported proposals. Finally, for such a transformation to occur, participants must not only aim to persuade their interlocutors, but must also be willing to listen to others and to revise their positions in light of their contributions.⁴⁰

Importantly, while participants should strive to persuade each other, culminating in a consensus is not a necessary mark of good public deliberation. Although earlier deliberative democrats suggested that successful deliberation tends towards a unanimous agreement,⁴¹ this view is now widely rejected. Instead, deliberation might well culminate in a majoritarian vote, which is itself followed by more deliberation in light of the information made available by implementing the vote's result. There are two reasons for rejecting the requirement of consensus. First, achieving a consensus often seems infeasible in profoundly divided settings. Second, having some participants register disagreement with policies that are ultimately adopted can be positively useful: dissidents have an important role to play following votes in continuing to publicly scrutinise chosen policies.⁴²

Note, finally, that although deliberation-centred democracy requires active participation by ordinary citizens, it is not synonymous with direct democracy. For most deliberative democrats, inclusive deliberation is intimately bound up with political representation. Citizens should deliberate with each other regarding whom they should elect, argue about whether their representatives are doing their jobs, contact those

⁴⁰ For an overview, see Mansbridge et al (2010). Public deliberation is 'public' at least in the sense that it is about public or political affairs. Often, it is also public in that it is visible to and addressed to the demos generally. I say 'often' because, as Chapter 2 will discuss, the systemic understanding of public deliberation recognises the importance of having more closed deliberative spaces—e.g., consciousness-raising groups—provided that the deliberation occurring in these spaces aims eventually to feed into more open deliberation—e.g., media communications, parliamentary debates.

⁴¹ E.g., Joshua Cohen (1989, 22), Habermas (1996).

⁴² E.g., Anderson (2006, 15-17), Mansbridge et al (2010, 66-67), Young (2000, 40-44).

representatives to demand justifications, and publicly criticise representatives who break electoral pledges.⁴³

With this minimal account of deliberative democracy in mind, consider why the Democratic Solution constitutes a *prima facie* promising dissolution of Section 2's dilemma. On the one hand, unlike the Realistic Solution, the Democratic Solution preserves the aim of realising justice. On the other, inclusive deliberation seems instrumentally and intrinsically valuable for justice, in a way that helps it address the Moral Imposition Solution's three problems.

First, deliberative democracy is instrumental to justice. Inclusive deliberative processes, it is often argued, pool diverse social perspectives. In doing so, they help us identify just policies in contexts where relevant information is dispersed across a heterogeneous population.⁴⁴ Therefore, deliberative democracy seems well-adapted to the *epistemic* difficulties that stand in the way of just policy-making.

Second, democratic procedures are also intrinsically valuable for justice, as they constitute a bulwark against *domination* (which, recall, I am taking to be *prima facie* unjust). Inclusive reason-giving processes allow citizens to vocally challenge political power and thereby make it more responsive to their concerns. Thus, these processes constitute a check on political power, which prevents such power from being exercised in a purely arbitrary fashion. Since domination just is arbitrary or unchecked rule, this means that deliberative democratic procedures are constitutive components of non-domination.⁴⁵

Finally, and related to this last point, weaker parties can use public reason-giving processes to make their concerns visible to the powerful and to hold them accountable. This makes the Democratic Solution more sensitive to non-ideal *contexts of power*, where those who know what is just often lack the power to forcibly impose it: since political change is driven by reason-giving rather than brute force, weak parties with valuable insights into what justice requires stand a better chance of effecting just change.

The Democratic Solution therefore seems an appealing response to the Promotion Problem. Nevertheless, while I take these considerations to constitute compelling

⁴³ E.g., Urbinati (2000), Young (2000, ch. 4), Mansbridge (2003).

⁴⁴ On deliberation's epistemic value, see Young (2000, 115-120), Anderson (2006, 10-15), and Landemore (2013, ch. 4). Chapters 2 and 4 expand on the epistemic value of democratic deliberation.

⁴⁵ On deliberation and non-domination, see Forst (2014, 201), Young (2000, 32-33), Pettit (2012, chs. 4-5), and Anderson (2010, ch. 5). Chapters 2 and 3 expand on how democratic deliberation upholds non-domination.

grounds for endorsing this solution, the remainder of this chapter aims to qualify this optimism.

3.2. The Ambivalence of Democracy

The picture of the Democratic Solution outlined above suggests that deliberative democracy produces a virtuous circle of justice. Inclusive deliberation pools different perspectives on the unjust social context. This enables weaker parties to make themselves heard, and to hold the powerful accountable. Greater justice results, which in turn facilitates the democratic process by further empowering weaker parties and increasing the degree to which powerful parties are accountable to them.

However, this picture presupposes that weaker participants *can* make themselves heard through public debate. But what if background circumstances are so unjustly unequal that this is false? Examples are easy to come by. Historically, women and people of colour have routinely been unjustly denied formal participatory rights. But even individuals who are granted formal participatory rights are regularly silenced. In the US, vast socioeconomic inequalities enable wealthy Americans to dominate public discussion, thereby drowning out the concerns of more vulnerable groups. Similarly, when disempowered groups have persistently been vilified in public discourse, their social standing might be so low that no one takes their contributions seriously. Last, but not least, participants' claims are sometimes discounted or derided, for example if they are inarticulate or emotional, because they violate expectations regarding what counts as proper public discourse.

When these unjust exclusions exist to a very significant degree, powerful actors may not need to be responsive to weaker parties. If weaker parties cannot audibly voice their complaints, then, contra 3.1's discussion, the powerful can simply ignore them. In fact, powerful actors may conceivably use public deliberation to amplify their power. Engaging in public deliberation that is biased in their favour by background injustices might enable powerful actors to promote their preferred policies with an air of legitimacy they would have otherwise lacked. As Young summarises, then, if we start from very unjust circumstances,

we have a different circle: Where there are structural inequalities of wealth and power, formally democratic procedures are likely to reinforce them, because privileged people are able to marginalize the voices and issues of those less privileged.⁴⁶

The point is that the Democratic Solution presupposes that a certain threshold of background justice has been achieved. When it has not, formally democratic procedures may not only fail to promote justice, but may serve to amplify existing injustices.⁴⁷ Put differently, there is a level of background injustice at which public deliberation will produce not a virtuous circle of justice, but a vicious circle of injustice.

What are the background threshold conditions for democracy to be capable of combatting rather than amplifying injustice? As the above examples suggest, they can roughly be classified into three categories.⁴⁸ Democracy's justice-promoting powers depend on:

Formal Inclusion: The extent to which those who should be part of the demos are actually granted formal participatory rights, such as voting rights.

Informal Inclusion—Material and Symbolic Conditions: The extent to which members of the demos have the material and symbolic resources needed to participate meaningfully in political life, such as wealth, cultural standing, education, leisure.

Informal Inclusion—Norms of Public Deliberation: The extent to which the norms governing what counts as appropriate public discourse allow important concerns to be expressed by all members of the demos.

Thus, democracy has an ambivalent relation to justice. Its ability to eliminate rather than amplify injustice depends on how well it includes members of the demos, which in turn depends on whether various kinds of background conditions are satisfied.

⁴⁶ Young (2000, 34).

⁴⁷ On democracy's ambivalent relation to justice, see also Forst (2012, 198-200), Shapiro (2003, 30-33), and Fung (2005, 401).

⁴⁸ For discussion, see Young (2000, 52-53) and Fraser (2009, 16-21).

Therefore, the following question arises: how should these threshold conditions, or basic justice, be realised or brought into being? The next section considers how democratic theorists can and have answered this question. These answers, I will suggest, ultimately replicate the dilemma we started with.

3.3. Reproducing the Dilemma: Three Common Responses

a) The Constitutionalist Response

One common response recommends that deliberative democracy's threshold conditions be forcibly realised independently of democratic decision-making.⁴⁹ Moreover, because these threshold conditions are presuppositions for the moral desirability of democratic decision-making, this view suggests that threshold conditions should be enforced *against* democratic decision-making, should the decisions made through inclusive deliberation and majoritarian voting conflict with these conditions.

This last point is most notably illustrated by philosophers who, like Cécile Fabre, endorse a form of constitutionalism. Here, I understand constitutionalism broadly as the claim that democratic decision-making processes should be constrained by a constitution. A constitution, on this view, enshrines certain basic rights and duties, including rights to participate politically on certain terms, and (if this is what effective democratic participation presupposes) certain socioeconomic rights. Its strictures must be enforced by unelected independent bodies, like judicial courts, even when the demos or its elected representatives produce a contrary decision.⁵⁰

Unfortunately, this 'constitutionalist' response seems like a version of the Moral Imposition Solution, applied restrictively to the realisation of democracy's threshold conditions.⁵¹ Accordingly, it encounters the familiar difficulties we outlined in 2.1. First,

⁴⁹ E.g., Fung (2005, 401-404).

⁵⁰ Fabre (2000a), for instance, holds that basic rights, including the preconditions of proper democratic functioning (96-97), should be upheld by a coercively-enforced constitution (86-87), and that this enforcement should be maintained even against opposing majoritarian decisions (83-85). See also Blake (2005, 232-235), Rawls (1997, 766), and Pettit (2012, 220-225). Now, most constitutionalists recognise that constitutions *can* be changed through democratic decision-making, provided certain demanding conditions are met, such as supermajority voting. E.g., Fabre (2000a, 96), Rawls (1997, 772), Pettit (2012, 238). What this shows is that constitutionalism comes in degrees (varying with how demanding the amendment conditions are). Strictly speaking, then, my claim in this section is that *insofar as* constitutionalists empower an unelected minority charged with forcibly enforcing basic or threshold justice against democratic decision-making, they replicate many of the Moral Imposition Solution's difficulties.

⁵¹ See Walzer (1981, 384-388) for a similar claim.

having unelected agents impose threshold conditions, sometimes against the democratic procedures instituted to check political power, seems at least *pro tanto* in tension with non-domination.⁵² So, if we could find an alternative strategy that avoids dominating individuals when bringing about democracy's threshold conditions, this would be morally preferable.

This first criticism of the constitutionalist response might seem wrongheaded. A constitution typically prescribes many checks on how political power can be exercised. So enforcing a constitution, one might think, is subject to checks and therefore does not constitute arbitrary rule, or domination.⁵³ One problem with this response is that it presupposes the *outcome* of forcibly bringing about the constitution. Our question is how one should realise democracy's threshold conditions in the first place. *Before* the constitution enshrining democracy's threshold conditions and defining constraints on political power exists, non-democratic force used to bring it into being is not subject to those constraints. The French Revolution may have resulted in republican constraints on political power, but the revolution itself was not so constrained.

Admittedly, however, constitutions that aim to protect the conditions required for democracy to function well are already a political reality in many contemporary democracies. In the US and across Europe, legal constitutions protected by independent courts, like the US Supreme Court, are historically well-entrenched. Moreover, as Ian Shapiro notes, "written constitutions [...] with independent courts charged to enforce them, are near-universal features of the new democracies that have swept the world since the 1970s".⁵⁴ In these cases, enforcing the constitution might well be subject to various checks and therefore non-dominating.⁵⁵

Nonetheless, enforcing democracy's threshold conditions non-democratically encounters the other difficulties that plagued the Moral Imposition Solution, *even when* we are enforcing an existing constitution rather than bringing one into existence. To begin, like the Moral Imposition Solution, the constitutionalist response seems epistemically immodest. As we have seen, politically-relevant information tends to be dispersed throughout society, making it difficult for any one person or group to gather all the information that is relevant to just decision-making. This epistemic obstacle was

⁵² See also Waldron (1993, 28), though he uses the language of autonomy, not non-domination.

⁵³ Pettit (2012, 236-237).

⁵⁴ Shapiro (2016, 41).

⁵⁵ For Bellamy (2008, 168-172) and Shapiro (2012, 329-332), constitutionalist provisions are dominating even in such cases.

one motivation for making decisions via inclusive deliberative processes. Accordingly, forcibly imposing democracy's threshold conditions, independently of inclusive deliberative democratic processes, seems epistemically problematic.⁵⁶

To appreciate this, consider that what exactly is needed for citizens to be capable of participating meaningfully in democratic public deliberation is a highly complex and contested question. Even if we agree that basic justice requires that citizens be able to participate in democratic life as social and political equals (as I have been assuming), how much material redistribution, what kinds of symbolic recognition, and what specific norms of political discourse are required for this to be possible remains an extremely challenging question. Consequently, even in societies that already possess a constitution purporting to protect the conditions for adequate democratic participation, there is often extreme disagreement concerning how these constitutional protections should be interpreted. Now, the disparate experiences that are distributed across the demos seem directly relevant to determining what citizens need for adequate democratic participation. Hence, leaving the task of realising democracy's threshold conditions to non-democratic, unelected bodies seems (like the Moral Imposition Solution) epistemically insensitive to the social dispersion of information.

Accordingly, as Jeremy Waldron has argued, there is no shortage of court decisions interpreting constitutions that produced deeply unjust decisions. Consider a few US Supreme Court cases. Notoriously, the Supreme Court ruled in *Plessy v. Ferguson* that equality is compatible with racial segregation. In *Schenck v. United States*, it seriously curtailed freedom of speech by deciding that criticising conscription during World War I was similar to shouting fire in a crowded theatre. During World War II, it legitimised the indiscriminate ordering of Japanese Americans into internment camps. More recently, it drastically amplified the political repercussions of economic inequality by declaring, in *Buckley v. Valeo*, that limits on electoral spending were unconstitutional.⁵⁷ Examples like these support the epistemic argument against forcibly realising democracy's threshold conditions independently of democratic processes of inclusive deliberation—i.e., independently of deliberation which involves ordinary citizens

⁵⁶ For an epistemic critique of constitutionalism and judicial review, see Waldron (2006, 1376-1386). Fabre (2000a, 79-80) acknowledges this objection but brackets it and simply assumes that "the judiciary would do a good job of protecting rights".

⁵⁷ Waldron (2006, 1376-1378). See also Shapiro (2016, ch. 4) for discussion.

directly or via their elected representatives.⁵⁸ History is riddled with instances of unelected bodies like the Supreme Court making judgments that exacerbated existing injustices and took us further away from democracy's threshold conditions.

Finally, like the Moral Imposition Solution, advancing democracy's threshold conditions as the constitutionalist response suggests is problematically insensitive to non-ideal contexts of power. To see this, notice that we can interpret the aforementioned Supreme Court cases in two ways. On the above interpretation, they suggest that correctly ascertaining how to promote basic justice, or democracy's threshold conditions, is very difficult unless we pool different social perspectives in inclusive democratic processes of reason-giving. A second and weaker interpretation is the following. Even if one thinks that a particular group is epistemically privileged concerning how to achieve basic justice, these cases remind us that, in non-ideal contexts, those who are powerful and those who are right about what basic justice demands seldom coincide. And it is the powerful who are in a position to enforce their views. Even if black Americans knew that racial segregation was clearly incompatible with democracy's proper functioning, they were not in a position successfully to impose their views. The all-white Supreme Court was.

Thus, on the second interpretation, these cases manifest the problem that in contexts of disagreement, recommending the non-democratic imposition of basic justice begs the question of who should do the imposing. So, against a background of unjust inequalities where the powerful may have incorrect views regarding what basic justice and the proper functioning of democracy require, applying this recommendation may well be inimical to realising democracy's threshold conditions.

In sum, the three challenges to the Moral Imposition Solution resurface to a significant degree against the popular view that we should simply forcibly impose justice up to the threshold where democracy becomes justice-promoting.

b) The Agonistic Response

In recognition of these problems, the Democratic Solution's proponent might reject the idea of imposing democracy's threshold conditions through non-democratic means.

⁵⁸ This qualification is important: while the Supreme Court may deliberate—though Waldron (2006, 1382-1386) expresses doubts about this—it is not an instance of *democratic* deliberation. Indeed, its deliberation does not include ordinary citizens or their elected representatives. For discussion, see Dryzek (2000, 15-16).

But if we do not impose these conditions, how *do* we realise them? One prominent family of democratic theories, agonistic democracy, responds by denying the question's presupposition. To understand what I mean by this, consider two central features of agonistic democracy, as expounded by Chantal Mouffe. First, agonistic democrats criticise classical formulations of deliberative democracy for abstracting from real-world power relations. Power and antagonism, Mouffe claims, are ineradicable.⁵⁹ So, a realistic conception of democracy should not abstract from them, but should rather take its starting point within existing antagonistic power relations. More specifically, "the aim of democratic politics is to transform an 'antagonism' into an 'agonism'".⁶⁰ Whereas antagonism describes the relation between enemies, who are hostile towards each other, agonism is the relation between adversaries, who struggle with each other but do not question each other's right to participate. For our purposes, what is significant is that by portraying democratic decision-making as embedded within antagonistic power relations, Mouffe seems to be *denying* that we should try to realise threshold conditions, where real-world power imbalances are significantly reduced, prior to democratic decision-making.

Relatedly, agonistic democracy tends to oppose restrictions on the content and form of reasons raised in public deliberation. Mouffe, for instance, rejects a conception of democracy that "eliminate[s] passions" from public discourse, and advocates an inclusive struggle between radically different ideas. Alongside other agonistic democrats, then, she sometimes suggests that even extreme ideas should be admissible in democratic discourse.⁶¹ In summary, agonistic democracy interprets democracy as taking place not in an idealised context where threshold conditions of basic justice are already satisfied but here and now; accordingly, it strives to accommodate the immense diversity of viewpoints that characterises actual societies.

While there is much to admire in the aim of making the democratic ideal more relevant to non-ideal circumstances, this position seems problematic in our context. If no effort is made to remedy existing antagonisms and power imbalances prior to democratic decision-making, and, relatedly, no effort to block hateful forms of extreme speech, why think that democracy will help us eliminate existing injustices? After all, the problem we raised in 3.2 was that against a background of severe injustice, formally

⁵⁹ Mouffe (1999, 752). See also Young (2000, 49-51), Honig (1996, 258), and Schaap (2006, 257, 264).

⁶⁰ Mouffe (1999, 755).

⁶¹ Ibid. See also Young (2000, 24, 49-51), Tully (2005, 206-208), and Schaap (2006, 269).

democratic decision-making might well generate a vicious circle of injustice instead. Sensing this problem, Young, who sometimes identifies as an agonistic democrat,⁶² asserts that democratic theorists

believe that the circle [of injustice] can be broken. In formally democratic societies with serious injustices it must be possible to promote social changes towards greater justice through democratic means⁶³

But what does Young mean by this “must”? Unless a reason is given for thinking that, contra the considerations raised in 3.2, formally democratic processes are bound to be justice-promoting, this seems worryingly like wishful thinking.

Mouffe might reply that I am misconstruing agonistic democracy. Perhaps her view is not that democratic politics *is* the transition from antagonism to agonism, but rather that it *results* from this transition. This would make sense of the kinds of processes Mouffe takes as paradigmatically democratic, since these processes do not seem to involve unfettered power or enmity. As mentioned above, Mouffe depicts democratic communication as a struggle between adversaries who profoundly disagree but nevertheless acknowledge each other’s right to defend their ideas. She later adds that participants in inclusive democratic communication are united by a “shared adhesion to the ethico-political principles of democracy”.⁶⁴

Although this second understanding of agonistic democracy yields a more appealing picture of democracy, it is unclear how it is supposed to solve our problem. It clearly *does* presuppose the prior realisation of background conditions. Most obviously, despite sometimes suggesting the opposite, Mouffe ultimately excludes some appeals from public communication: in particular, participants must not challenge the “ethico-political principles of democracy”. Even for agonistic democrats like Mouffe, then, some abhorrent extreme views should be excluded from democratic processes.⁶⁵ Consequently, the question resurfaces: how do we realise these background conditions which effective democratic public communication presupposes? On the second

⁶² Young (2000, 49-51).

⁶³ Ibid., 35.

⁶⁴ Mouffe (1999, 755). See also Young (2000, 25).

⁶⁵ On this point, see Knops (2007, 116-118).

understanding of agonistic democracy, agonistic democrats have not answered our problem.

Crucially, a new parallel with Section 2 arises: the difficulties agonistic democracy encounters parallel those that obstructed the Realistic Solution in 2.2. The Realistic Solution, recall, either effectively abandoned the aim of promoting justice, or reproduced the Moral Imposition Solution's difficulties. Analogously, the agonistic response either effectively abandons the aim of realising democracy's threshold conditions, or reproduces the problem of how we should impose these conditions. In this light, the agonistic response is to the problem of realising democracy's threshold conditions what the Realistic Solution is to the Promotion Problem—that is, the broader problem of how we should go about realising justice in non-ideal conditions. Both solutions offer a conception of politics that takes as given unjust power relations, and that pursues an aim more modest than fully eliminating them. While the Realistic Solution aims for peace rather than justice, the agonistic solution suggests transforming rather than eliminating power relations within democratic politics. Both, in consequence, are hostage to existing injustices in a way that yields an unappealing picture of politics. And finally, both tend to shy away from this upshot, thereby reproducing the strategy they criticised in their opponents.

c) The Meta-Democratic Response

A final possible response does not deny that democracy's proper functioning presupposes background conditions of basic justice, but argues that the best way to realise these conditions is through democratic deliberation on democracy itself. On this view, inclusive deliberation takes as its object the existing deliberative political system and asks how it should be improved. This approach is particularly popular among discourse-theorists, such as Habermas, Dryzek, Fraser, and Benhabib.⁶⁶

To illustrate this 'meta-democratic' strategy, consider Benhabib's discussion of who constitutes the demos. Benhabib declares that "membership in the *demos*, no matter how arbitrary initially, can be self-reflexively adjusted only by the *demos* themselves".⁶⁷ Put differently, deliberation should start from the currently-recognised demos, however

⁶⁶ Habermas (2001, 774); Dryzek (2010, 12); Fraser (2009, 27-29); Benhabib (2006, ch. 2).

⁶⁷ Benhabib (2011, 165).

arbitrarily or unjustly it was formed. In turn, that demos's reflexive deliberation will produce a more just delineation of the demos.⁶⁸ As evidence, Benhabib notes how public deliberation in Germany led to new citizenship laws in 2000, which made acquiring German citizenship easier for residents—a change which, she suggests, resulted in a fuller realisation of democracy.

The problem with this view is that its advocates rarely distinguish between meta-democracy before the democratic threshold conditions have been achieved and meta-democracy afterwards. That it is meant to apply to both is clear, for instance, from Dryzek's affirmation that meta-democratic deliberation should not simply occur in well-functioning democracies,⁶⁹ as well as from Benhabib's above statement that meta-deliberation starts from current conditions, "no matter how arbitrary" they are. Failing to distinguish between these two kinds of cases makes the meta-democratic strategy seem eminently plausible. Indeed, as Benhabib's example shows, it is undeniable that democratic deliberation sometimes improves itself.

However, when we turn to the specific question of how democracy's threshold conditions should be promoted, this seems an unpromising solution. I argued in 3.2 that before the threshold conditions are realised, democratic deliberation may amplify injustice rather than remedy it. Therefore, for the meta-democratic strategy to be promising, meta-democrats *either* need to refute the argument offered in 3.2; *or* they need to demonstrate that there is a disanalogy between deliberation on democracy itself and normal deliberation such that the effectiveness of the former does not presuppose that a basic threshold of justice has already been attained. In the absence of arguments for either claim, the meta-democratic strategy is unconvincing.

Moreover, there are reasons to doubt that a successful argument for either claim is forthcoming. In contexts marked by deep background inequalities, democratic decision-making *does* sometimes damage democracy itself. In the US, initiatives aimed at reducing the vast socioeconomic inequalities that stand in the way of genuine political equality have encountered immense legislative resistance. Instead, tax cuts enacted under the Bush and Trump administrations seem likely to exacerbate these democracy-

⁶⁸ Benhabib (2006, 451).

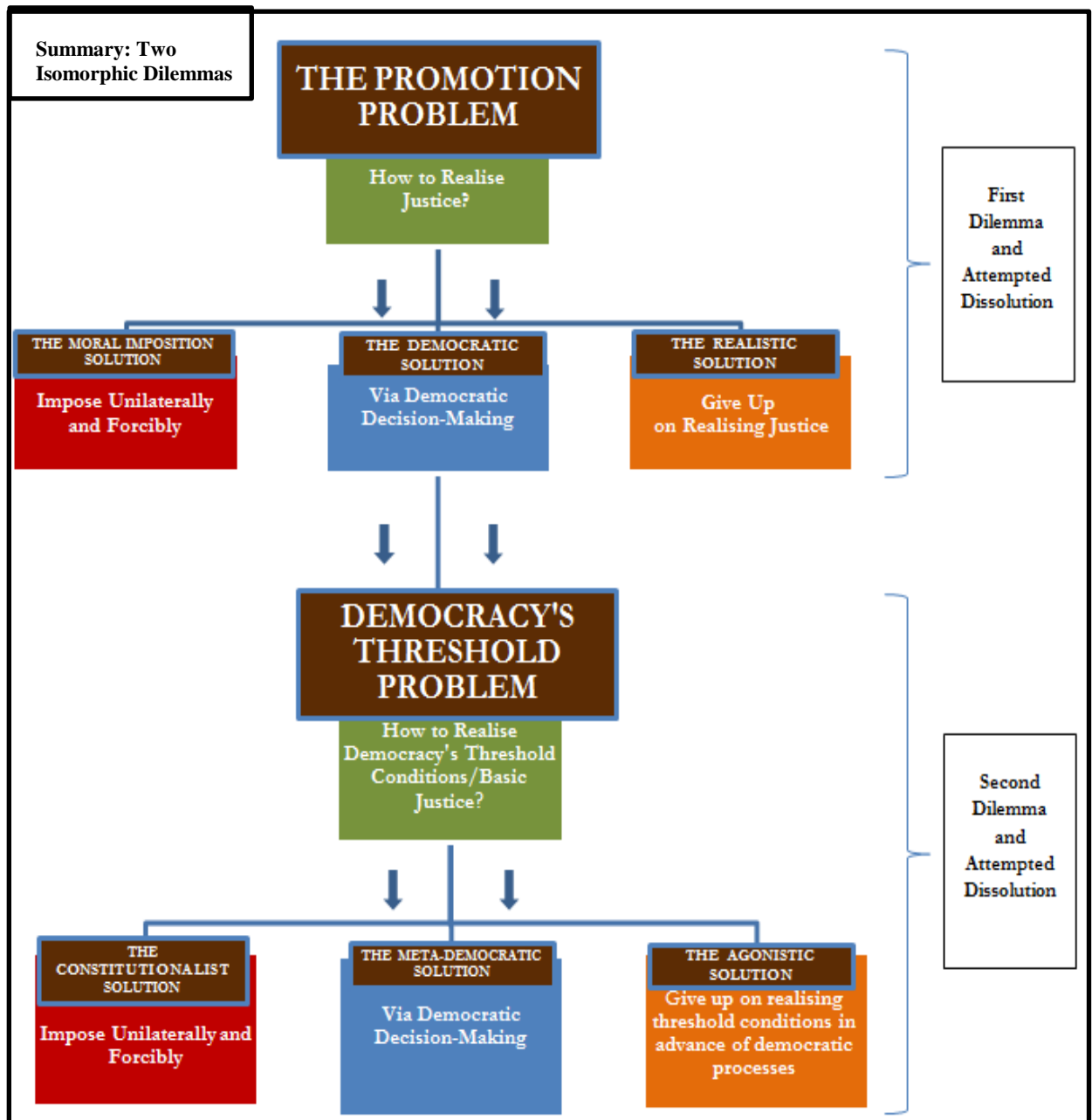
⁶⁹ Dryzek (2010, 146-147).

impeding inequalities.⁷⁰ Meta-democracy, then, simply does not inevitably improve democracy.

In light of these difficulties, a third and final parallel appears: the popular meta-democratic strategy is to the question of how we should promote democracy's threshold conditions what the Democratic Solution is to the Promotion Problem. While the latter recommends promoting justice through democratic decision-making, the former asserts that we should promote democracy's threshold conditions, or basic justice, through democratic decision-making. Predictably, then, they are threatened by the same ambivalence. Amidst vast background injustices, both risk promoting injustice rather than justice.

Having surveyed these three common democratic responses, let us take stock. The present section suggests that the Democratic Solution encounters a dilemma when addressing the question of how we should realise the threshold conditions needed for democracy to be an effective tool for promoting justice. If the answer is 'through non-democratic forcible imposition', the challenges familiar from the Moral Imposition Solution resurface (3.3.a). If, to avoid these difficulties, we follow agonistic democrats in resisting the presupposition that we should eliminate significant unjust power imbalances prior to resorting to democratic decision-making, then it is unclear why democracy will advance justice rather than injustice (3.3.b). Finally, an attempt to go between the horns of this democratic dilemma by suggesting that the threshold conditions should be realised democratically unsurprisingly recreates the Democratic Solution's ambivalent relation to justice (3.3.c). As I have stressed throughout this section, this democratic dilemma closely parallels the dilemma that motivated the turn to democracy (see diagram below). Hence, an uncomfortable result arises: the dilemma that the Democratic Solution was introduced to dissolve resurfaces internally to democratic theory.

⁷⁰ Scott and Chang (2017).



4. Conclusion: A Diagnosis

I began by outlining a dilemma for the Promotion Problem, the problem of how we should go about realising justice in non-ideal contexts characterised, among other things, by deep ethical divisions. On the one hand, recommending the non-democratic forcible imposition of justice evokes the ominous figure of Plato's philosopher king, along with its familiar problems: its problematic relation to domination, its epistemic immodesty, and, perhaps most importantly, its insensitivity to non-ideal contexts of

power. ‘Realistic’ responses that adopt a more modest aim for politics than realising justice, on the other hand, seem *too* sensitive to existing power imbalances. By taking too much of the background context as given, they allow politics and its aims to be taken hostage by pre-existing injustices.

4.1. The Inevitable Cost of the Democratic Solution

In this context, I examined the view that the Democratic Solution can navigate between the horns of this dilemma. My first conclusion is that although there are good grounds for thinking that the Democratic Solution is the most promising way of promoting justice, it encounters a dilemma similar to the one it aimed to avoid. Unless democratic deliberation presupposes exercises of non-democratic moral imposition, it risks being co-opted by existing unjust power imbalances and becoming complicit in their reproduction.

So, if we want the Democratic Solution to effectively promote justice, it seems we cannot avoid the conclusion that some non-democratic forcible imposition is required in establishing democracy’s threshold conditions—i.e., the minimal conditions that must obtain for democratic deliberation to advance justice rather than injustice. In this respect, the constitutionalist approach to democracy is correct. Consequently, the Democratic Solution cannot fully escape the costs associated with the Moral Imposition Solution.

4.2. An Unhappy Conclusion

What should we make of this initial conclusion? We can discern two types of reactions. The first, illustrated by agonistic democrats and meta-democrats, consists in seeking, unsuccessfully, an alternative way of realising democracy’s threshold conditions. Their vigorous denial that democracy must run afoul of the Moral Imposition Solution’s difficulties tacitly acknowledges how seriously problematic this conclusion is.

But not all political philosophers feel this way. In particular, constitutionalist democrats who explicitly bite the bullet by endorsing the non-democratic moral imposition of democracy’s threshold conditions might and sometimes do reply as follows: ‘Yes, the threshold conditions for deliberation to effectively promote justice

must be realised by non-democratic means. So what? The Democratic Solution is still better than the alternatives canvassed. It is better than the Realistic Solution because it does not abandon the aim of promoting justice. And it is better than the Moral Imposition Solution because, although there are indeed costs to undemocratically imposing justice, the Democratic Solution requires such moral imposition to a lesser extent. It only requires moral imposition up to the point where, although some injustice remains, conditions are sufficiently good for inclusive deliberation to produce a virtuous circle of justice. Hence, although it reproduces the dilemma we began with, it reproduces it in an attenuated form. The Democratic Solution's problems should therefore not worry us.⁷¹

My second conclusion is that, contra this second response, we should be deeply unhappy about this chapter's primary result. The present chapter has shown not just that the Democratic Solution inevitably incurs certain costs, but also that these costs are serious. Indeed, I have argued that the difficulties the Democratic Solution encounters are similar in kind—if not quite in degree—as those faced by the Moral Imposition Solution. Given how crude and undesirable this solution is—as I noted in 2.1, contemporary philosophers widely reject it—it is a profoundly worrying result that the Democratic Solution faces the same (albeit somewhat attenuated) problems. Accordingly, as it stands, even if the Democratic Solution is the lesser evil among the alternatives canvassed, the evil it involves seems very significant in absolute terms.

4.3. The Challenge Remaining

Even so, this last observation might give rise to a second 'so what?' question. Lamenting the problems that afflict the best solution to the Promotion Problem seems futile if nothing can be done about them.

In fact, however, there *is* something we can do about the unhappy result we have arrived at. The Democratic Solution runs afoul of the Moral Imposition Solution's difficulties *to some extent*. To what extent it does depends on what the threshold is for democracy to be an effective tool for promoting justice. Different theories of democracy place that threshold in different places. More idealising theories of democracy, like Habermas's, Cohen's, or Estlund's, outline an ideal deliberative

⁷¹ For a similar response, see Fabre (2000a, 96-98).

situation which abstracts almost entirely from the injustices and power relations that characterise actual politics.⁷² Such theories, then, set the threshold at which valuable deliberative democratic processes begin very high. By contrast, agonistic democrats place the threshold conditions much lower, if they recognise any such conditions at all. The lower the threshold, the less the Democratic Solution presupposes non-democratic moral imposition.

In this light, the challenge ahead should be to rethink deliberative democracy for non-ideal circumstances. More precisely, the rest of this inquiry seeks to address the following question: what would deliberative democracy have to look like for its characteristic discursive situation to be as close to actual imperfect conditions as possible, while remaining an effective process for realising justice in those conditions? Exploring this question will allow us to develop a deliberative theory of democracy that minimises the costs of moral imposition, without giving up on promoting justice. Put differently, it will help us achieve the best balance between the need to ensure that democratic processes are effective at realising justice and the need to avoid as far as possible the important costs associated with moral imposition. In this way, we can turn the ‘lesser evil’ into the least possible evil. Following the pioneering work of Iris Marion Young,⁷³ then, what follows is an attempt to articulate a vision of democratic justice for non-ideal conditions.

More specifically, I will focus on what democratic public deliberation would have to be like to constitute an effective tool for advancing justice in non-ideal conditions. Hence, my primary emphasis will be on what norms should regulate the content and form of inclusive public deliberation (Chapters 2-4), and on the extent to which the attitudinal norms presupposed by such deliberation are compatible with existing material and symbolic background conditions (Chapters 5-6). The question of who exactly should be granted formal participatory rights is one that I set aside here.

Before proceeding, two observations are in order. First, as I have stressed, showing that the Democratic Solution encounters a dilemma that is structurally similar to the one that prompted its introduction is important, because it motivates rethinking democratic deliberation. But bearing this initial dilemma in mind also matters because it

⁷² Habermas (1996, 305-306); Joshua Cohen (1989, 21-23); Estlund (2008, 18-20). Bächtiger et al (2010) distinguish more broadly between Type I and Type II approaches to deliberative democracy, with the former placing the threshold higher than the latter.

⁷³ Young (2000). See also Williams (2000, 144-146), Chambers (2009), and Anderson (2010, ch. 5).

offers useful conceptual lenses for understanding what is going on within democratic theory. We have already seen in 3.3 that we can helpfully classify prominent theories of democracy based on what horn of the original dilemma they tend to run afoul of. For similar reasons, keeping this dilemma in sight will help us orient our thinking and avoid certain pitfalls while developing a non-ideal theory of deliberation. Indeed, when discussing specific norms and presuppositions of democratic deliberation, we can expect the horns of the dilemma to resurface in more localised forms.

Second, it is worth pausing to consider the place of philosophy in this project. Determining under what conditions inclusive deliberation is justice-promoting rather than injustice-promoting is partly an empirical task. That is, it depends partly on social scientists' findings concerning the consequences of different discursive norms. But philosophy also has an essential role to play. To begin, suppose we have a model of democratic deliberation designed for ideal conditions. Finding out whether its recommendations remain desirable in non-ideal conditions requires knowing what the necessary preconditions of those recommendations are. And, as illustrated by the rich theoretical literature on trust and its relation to democracy (which Chapter 5 will consider) articulating and conceptualising such preconditions is a deeply philosophical task.

Furthermore, the conceptual and normative work performed by philosophers helps appreciate why non-ideal circumstances are sometimes part of what makes democratic discursive processes valuable in the first place. Take contexts where some citizens express degrading views about others. To determine why the counterspeech supplied through democratic public discourse is a valuable response to such utterances, we need to understand what kind of harm is produced by degrading utterances, and how democratic counterspeech relates to that harm. As Chapter 3 will show, philosophers' extensive conceptual work on language is directly relevant to this aim.

More generally, then, political philosophy contributes to answering the present question by offering a conceptual framework that facilitates and enriches subsequent empirical investigations into precisely what deliberative institutions should be implemented. With these background motivations and considerations in mind, let us now turn to developing an account of non-ideal democratic public deliberation.

PART 2
NORMS OF DEMOCRATIC
PUBLIC DELIBERATION

CHAPTER 2

SHARED REASONS AND THE EXCLUSIVENESS OF PUBLIC DISCOURSE

1. Introduction

In Chapter 1, I considered the attractive view that deliberative democratic processes are the most appropriate means of advancing justice in non-ideal conditions marked, *inter alia*, by significant ethical disagreement and pervasive inequalities. Deliberative democracy, recall, is a conception of democracy that gives a central role to inclusive public exchanges of reasons. Accordingly, the political equality it strives for involves giving citizens not only equal voting rights, but also an equal opportunity to participate, directly or indirectly, in public discussion.

Such democratic decision-making, I argued, constitutes an unpromising tool for promoting justice unless certain background conditions, democracy's 'threshold conditions', have already been achieved. The effectiveness of democratic processes at realising justice depends on several kinds of background conditions, including: to what extent those who should be in the demos are actually granted formal citizenship rights; how vast material and symbolic inequalities are; and what norms regulate the content and form of public deliberation.

Finally, I conceded that these threshold conditions may sometimes have to be realised non-democratically. Because this comes with serious costs, democratic theorists should be concerned with the following question: what would deliberative democracy have to look like for its 'threshold' deliberative situation to be as close to actual non-ideal conditions as possible, while remaining an effective tool for promoting justice? Answering this question should yield a vision of democracy that minimises the costs associated with non-democratically realising democracy's threshold conditions, without abandoning the aim of advancing justice.

In this chapter, and the following two, I examine this question in relation to the third kind of threshold condition. What rules should regulate the content and form of public deliberation in non-ideal conditions? The present chapter examines one of the

most widely defended discursive norms, the Shared Reasons Constraint, according to which participants should appeal only to reasons that are shared.¹

I will establish three claims. First, although there are powerful grounds, stemming from the value of non-domination, for upholding the Shared Reasons Constraint, recent work in social epistemology indicates that even the weaker variants of this constraint risk regulating the content of public discourse in an overly restrictive way. That is, they risk excluding too many considerations from public deliberation (Section 2). Second, the three most popular strategies put forward by democratic theorists for offsetting this ‘overexclusiveness’ encounter serious difficulties, both individually and taken together (Section 3). Lastly, I propose that we can mitigate these difficulties by giving a greater role to non-argumentative and emotionally-charged forms of speech, like narrative, in formal and informal arenas of democratic public deliberation (Section 4).

In establishing these conclusions, I am departing from typical non-ideal theories of democratic deliberation in an important way. While the Shared Reasons Constraint has been a staple of theories that articulate norms for public deliberation occurring under more or less idealised conditions (as Rawlsian public reason theory notably does²), non-ideal democratic theorists have more commonly been wary of this constraint.³ Part of what is distinctive in my approach, then, is the suggestion that there are powerful reasons to try to maintain a Shared Reasons Constraint in non-ideal conditions. The challenge then becomes how we can do so while minimising the extent to which this constraint is overly exclusive. That is what the present chapter investigates.

¹ Following Scanlon (1998, 18), I understand reasons as considerations that purport to count in favour of something.

² E.g., Rawls (1997), Quong (2011), and Estlund (2008).

³ E.g., Mouffe (1999, 755), Young (2000, 49), Talisse (2009, ch. 2), Schaap (2006, 263-264), Waldron (1999, ch. 7), and Dryzek (2010, 86-93). See note 67, this chapter, for discussion of Dryzek’s position. Now, some non-ideal deliberative theorists do explicitly defend the Shared Reasons Constraint. E.g., Williams (2000, 130), Gutmann and Thompson (1996, ch. 2), and Krause (2008, ch. 5). However, Williams’s discussion of non-ideal public deliberation is largely programmatic. As for other non-ideal theorists who embrace the Shared Reasons Constraint, such as Krause, Sections 3 and 4 will engage with their more specific arguments and positions.

2. Sharing Reasons: A Defence and an Objection

2.1. The Shared Reasons Constraint

Most political philosophers agree that some constraints should regulate the content of publicly-admissible considerations. But disagreements abound regarding what those constraints should be.⁴ Here, I wish to examine one of the most widely defended constraints on the content of public deliberation. Consider an initial statement:

The Shared Reasons Constraint: In public deliberation, one must appeal to reasons that are shared by all participants when justifying one's claims.

How exactly the various terms of this formula should be interpreted has been the subject of much debate. One source of controversy concerns the constituency of the Shared Reasons Constraint—i.e., who exactly is included under ‘participants’. Another concerns the precise site of the Shared Reasons Constraint—i.e., which specific forums of public deliberation this norm is meant to govern. I bracket the issues of constituency and site until Sections 3.2 and 3.3, respectively. Here, let us focus on the terms ‘must’ and ‘shared’.

The first thing to notice is that the Shared Reasons Constraint should not be understood as a coercively-enforced legal norm. John Rawls, who refers to this constraint as the “duty of civility”, is explicit about this: “this duty [...] is an intrinsically moral duty. I emphasize that it is not a legal duty”.⁵ Thus, deliberators ‘must’ appeal to shared reasons in the sense that they are morally required to appeal to such considerations. But the moral reasons supporting this obligation do not warrant legally suppressing public utterances of non-shared reasons. Nonetheless, lighter and more informal types of conventional enforcement may be warranted. In a parliamentary setting, for instance, this might manifest itself in having the presiding officer verbally rebuke speakers who merely appeal to non-shared considerations.⁶

In what sense should publicly-admissible considerations be ‘shared’ by participants? According to the most influential interpretation, a reason is shared when it is *acceptable*

⁴ For an overview, see Mansbridge et al (2010).

⁵ Rawls (1997, 769).

⁶ Ibid.

to all—that is, when all can accept that reason or consideration as normatively forceful. This interpretation is prominent in Rawlsian public reason theory. For instance, in his thoroughgoing rearticulation of the Rawlsian ideal, Jonathan Quong affirms that, to comply with the duty of civility, one must “offer a reason [...] that will be acceptable to everyone in their capacity as free and equal citizens. Public reasons, on this Rawlsian approach, are thus *shared* reasons.”⁷ But this interpretation is also common among deliberative democrats working outside of the Rawlsian tradition. Famously, Amy Gutmann and Dennis Thompson enjoin deliberators to “appeal to reasons that are [...] mutually acceptable in content”.⁸ In a similar vein, Sharon Krause advocates “framing arguments in terms of the common concerns that underlie the political order”.⁹ Thus, “public reasons are constituted by the things we *all* care about”.¹⁰

A standard example of a non-shared reason, on this interpretation, is an appeal to a specific religious worldview. That the Ten Commandments denounce killing is not a widely acceptable reason for outlawing murder. Notice, however, what the acceptability interpretation does *not* say. Firstly, this interpretation does not entail that the claim one is defending on the basis of acceptable reasons must also be deemed acceptable by others. If I defend tax cuts for the wealthy by appealing to their well-being, what matters is that you accept the consideration *that well-being is valuable*. But you may nonetheless deny that this policy would best promote it. Furthermore, the need to appeal to acceptable considerations does not necessarily prohibit appealing to one’s self-interest rather than to the common good. I could, for instance, publicly oppose tax cuts by declaring that well-being is valuable, and that they hinder my well-being.¹¹

Importantly, however, some theorists interpret ‘shared’ far more weakly. On an increasingly popular interpretation, which Daniel Weinstock has notably advanced, a reason is shared when it is *intelligible* to all—that is, when everyone understands its meaning.¹² While an acceptable reason is also intelligible to those who accept it, the converse is false. I can be acquainted with the Bible and understand appeals to

⁷ Quong (2011, 261; see also 10-11, 263). See, similarly, Rawls (1997, 770-771).

⁸ Gutmann and Thompson (1996, 57). See also Mansbridge et al (2010, 66), Williams (2000, 130), Habermas (1996, 22-23), and Joshua Cohen (1996, 100). Bohman and Richardson (2009, 274) defend a related view, according to which shared reasons are mutually accepted.

⁹ Krause (2008, 156; see also 18, 112, 156-159).

¹⁰ Ibid., 151.

¹¹ Mansbridge et al (2010, 73-80).

¹² Weinstock (2017, 647). See also Dowding (2018, 244), Vallier (2016), and Gaus (2010, 279-283).

particular divine commands, but nevertheless reject the view that they have normative force.¹³

Intelligibility is not the only alternative to acceptability as an interpretation of ‘shared’. Deliberative democrats have developed a wide range of competing interpretations, many of which occupy an intermediate position between acceptability and intelligibility. Gabriele Badano and Matteo Bonotti, for example, defend *accessibility*, which they take to be less permissive than intelligibility and more permissive than acceptability.¹⁴

But this diversity is unproblematic for my present argument. In what follows, after suggesting that the Shared Reasons Constraint (in both its stronger and weaker variants) has considerable *prima facie* appeal (2.2), I aim to diagnose a problem for this norm: even its weakest interpretation—intelligibility—is in important respects overly exclusive (2.3). If this criticism holds for the weakest interpretation of ‘shared’, then it will also hold for intermediate alternatives.

2.2. A Justification for the Shared Reasons Constraint

A powerful justification for the Shared Reasons Constraint concerns the freedom of those subjected to political power. While this justification holds on various understandings of freedom, here I focus on freedom understood as non-domination.¹⁵ Domination, recall, is subjection to arbitrary or alien control. One is subjected to another’s alien control when that other has “the capacity, not subject to [one’s] direct or indirect check, to interfere in [one’s] choice”.¹⁶ By contrast, being subjected to power does not dominate when that power is checked, so that it must track one’s concerns or interests.

How does the Shared Reasons Constraint relate to freedom, thus understood? One suggestion is that unless A, who is subjected to coercive power, appeals to reasons that are shared by B, who exercises that power, B will remain unmoved by her demands. Thus, to make political power responsive to her concerns and interests, and hence non-

¹³ E.g., Rawls (1997, 771), Vallier (2016, 598-615).

¹⁴ Badano and Bonotti (manuscript). Furthermore, there is diversity *within* interpretations of intelligibility: Vallier’s (2016) definition of intelligibility is somewhat stronger than the folk definition I give above.

¹⁵ See also McBride (2007, 176-179). Rawls (1997, 799-800) and Enoch (2015, 138) accept the broader point that the Shared Reasons Constraint advances freedom or autonomy.

¹⁶ Lovett and Pettit (2009, 14).

arbitrary, A must appeal to considerations that are shared by those wielding power. In this vein, Krause observes that Martin Luther King's demands for racial equality "became convincing to white Americans [...] because they appealed (in a logical fashion) to things most Americans really cared about: security, opportunity, and dignity, [...] liberty and equality".¹⁷

Krause is right that appealing to shared considerations contributes to effective political contestation (and therefore non-domination). But the argument suggested above falls short of establishing that conclusion. As stated, it is compatible with the prudential recommendation that A appeal to values B recognises, independently of whether A herself shares them. To see why A and B should appeal to shared reasons, we must say more. Recall that democratic deliberation involves the exchange of justifications over time. Accordingly, when subjects make claims relating to an exercise of power, they typically do not do so in a vacuum. Rather, they are often responding to whatever justifications have been given for that power in an ongoing debate. Indeed, citizens often hold political agents accountable by poking holes in the justifications they have offered.

Now, one's ability to *respond* to a justification is substantially enhanced when that justification was made using a reason one shares. This is most obvious with intelligibility. If I do not understand the reason given to me, it is unclear how I could even begin to respond to it. If a theocratic government condemns me on the basis of its esoteric religion, and I do not comprehend many of this religion's cryptic concepts, my ability to contest and check that condemnation is severely impaired, and I do seem subjected to alien power.

But a similar observation holds with acceptability. Even when one understands the terms of the debate, it is not uncommon to feel incapable of responding to certain charges because one feels that the debate has been misframed, framed using reasons one does not accept. If I feel alienated from the reasons that have been advanced, it may be impossible to articulate the points I feel are genuinely important in terms of those considerations. Suppose a government rejects same-sex marriage on the grounds that it is unnatural. A political opponent may well experience difficulties arguing that same-sex unions are natural, and think that it is simply irrelevant whether they are. Naturalness, she might think, is not a normatively significant category. But this leaves

¹⁷ Krause (2008, 164).

her in an awkward position. Either she attempts to respond to the charge of unnaturalness, which is likely to distort the considerations she finds genuinely significant; or she does not, but then she leaves the justification that those in power find most compelling unanswered. Thus, because the power she is subjected to is justified in terms she profoundly rejects, she is unable to force it to track the considerations she deems fundamentally important.

Consequently, the point underpinning the Shared Reasons Constraint is not simply the prudential suggestion that *giving* reasons one's interlocutors understand and eventually accept helps make them responsive to one's interests. It is also the insight that when one *is given* reasons one understands and accepts as a justification for political power, one is better able to hold that power accountable and make it less dominating. To ensure that, in a debate where participants exchange reasons over time, all participants are given reasons they understand and accept, the debate must be framed in terms of shared reasons.

This freedom-based justification is defeasible: although there are powerful pro tanto reasons to embrace the Shared Reasons Constraint, these may be outweighed by countervailing considerations. Having provided a defeasible case for this norm, then, I will now introduce what I take to be strongest consideration against it.

2.3. The Overexclusiveness Problem

Perhaps the most daunting objection to the Shared Reasons Constraint is that it is overly exclusive, or rules out too many considerations. This is especially evident on the strong interpretation of 'shared', where shared reasons are acceptable to all participants. As many critics have noted, given non-ideal societies' deep divisions, it is likely that for most moral considerations, *someone* rejects them.¹⁸ Homophobes may find same-sex unions worthless and unnatural. White supremacists might judge that racial minorities are innately inferior. And some ideological doctrines, religious and non-religious, assign women to a subordinate role. It seems a deeply unwelcome result that, according to the strong Shared Reasons Constraint, we should therefore refrain from appealing to the values of loving same-sex relationships, of racial equality, and of gender parity as reasons for or against policies.

¹⁸ E.g., Williams (2000, 133-134), Talisse (2009, 52-56), Young (2000, 40-41), Enoch (2015, 118).

One might take this as grounds for reverting to the weak Shared Reasons Constraint, which only requires that reasons be intelligible to all. But what proponents and critics alike typically overlook¹⁹ is that overexclusiveness also plagues the weak Shared Reasons Constraint. Contemporary work in social epistemology brings this out particularly clearly. In recent years, social epistemologists such as Miranda Fricker, Kristie Dotson, Charles Mills, and Elizabeth Anderson have extensively explored and theorised the obstacles that prevent relatively disempowered groups from making important group-specific concerns intelligible to others.²⁰

According to these theorists, different social groups, who occupy different social positions based on their race, gender, class, and so on, systematically experience different constraints and enablements. Consequently, they may be radically ignorant of one another's experiences and perspectives. This is particularly true, social epistemologists suggest, with respect to what privileged groups know about less privileged groups. For example, Mills and Anderson have both argued that because of wide-ranging segregation between white and black Americans, whites typically encounter very different facets of American society than blacks do. In consequence, significant portions of blacks' daily lives remain invisible to whites.²¹

These deep experiential divides, Fricker argues, generate a distinctive epistemic problem for disadvantaged social groups. In highly unequal societies, disempowered groups typically have less influence over the creation and diffusion of social meanings. Consequently, the pool of conceptual resources a society has for conferring meaning on activities and experiences ends up being especially adapted to dominant groups' experiences. The upshot is that disadvantaged groups struggle to make sense of and to articulate their group-specific experiences: "some significant area of [their] social experience [is] obscured from collective understanding".²² Fricker calls this unjust deficit of intelligibility a "hermeneutical injustice".²³

In extreme cases of hermeneutical injustice, marginalised groups themselves lack the conceptual resources needed to make sense of their own experiences and concerns. Here, Fricker cites the case of sexual harassment. Before the concept 'sexual harassment' was developed in the 1970s, Fricker suggests, those experiencing it lacked

¹⁹ E.g., Vallier (2016, 602), Badano and Bonotti (manuscript).

²⁰ Fricker (2007); Dotson (2011); Mills (2007); Anderson (2010).

²¹ Mills (2007); Anderson (2010, 44-50).

²² Fricker (2007, 158).

²³ Ibid.

the categories needed to fully understand the wrong they were being subjected to. Their experiences had to be conceptualised as mere instances of ‘flirting’ or ‘joking’.²⁴ By contrast, in the more common case of hermeneutical injustice, although victims of hermeneutical injustice possess the concepts needed to make sense of their own experiences, some of their interlocutors do not.²⁵ When this is the case, requiring that deliberators appeal to mutually-intelligible considerations—as the weak Shared Reasons Constraint does—keeps victims of hermeneutical injustice from appealing to concepts that are needed adequately to express important considerations or concerns.

To clarify this, consider several examples social epistemologists have examined. Micro-aggressions, Dotson explains, are brief and commonplace slights and indignities directed at members of stigmatised groups, often automatically and unintentionally.²⁶ They include, for instance, a white person instinctively checking that she still has her wallet when a black person walks by. When these behaviours are systematic, and when they cumulatively tend to express a denigrating stereotype, they are reasonably experienced with indignation.²⁷ Now, to someone who never experiences such behaviours, it might be puzzling how very slight actions could *ever* amount to more than a trivial offense. In other words, someone in this position could fail to grasp the very concept of a micro-aggression. Just as many pre-Leibnizian mathematicians were unable to conceive of how a sum of infinitesimals could amount to a finite number, such a person might be unable to conceive of how many involuntary slights could cumulatively compose something as morally problematic as an aggression. Accordingly, members of privileged groups might (and sometimes do) think that denouncing behaviours as micro-aggressions is misguided, and that those who do so are simply overly sensitive.²⁸ In such instances, a public justification invoking the concept ‘micro-aggression’ is ruled out by even a weak Shared Reasons Constraint, because the concept is unintelligible to some participants.

Similarly, Fricker cites marital rape as an instance of hermeneutical injustice.²⁹ Many states did not recognise marital rape as a crime until very recently. Some still do not. Importantly, what made criminalising marital rape difficult was partly the concept’s

²⁴ Ibid., 152-153.

²⁵ Fricker (2013, 1319-1320).

²⁶ Dotson (2011, 246).

²⁷ See McTernan (forthcoming) for discussion of the injustice enacted by micro-aggressions.

²⁸ E.g., Lukianoff and Haidt (2015), Campbell and Manning (2015).

²⁹ Fricker (2007, 155).

perceived unintelligibility. To some, marital rape seemed a contradiction in terms, because rape presupposes a lack of consent, whereas marriage (as they conceived it) produces the relevant consent by definition. And indeed, even where marital rape *is* criminalised, public pronouncements denying the coherence of this concept remain common.³⁰ Here again, certain extremely significant considerations were not even weakly shared because they invoked concepts that were unintelligible to some.

Finally, and returning to same-sex marriage, the issue is not merely that same-sex relationships are not recognised as valuable, but also that what they *are* can be fundamentally misunderstood. Homosexuality, Fricker notes, was once classified by hermeneutical authorities as an illness or disorder. In the US, for instance, it remained in the Diagnostic and Statistical Manual of Mental Disorders (DSM) until 1973. And witness, even today, the tragically common amalgamation of homosexuality and paedophilia.³¹ In these cases, in addition to the value of same-sex relationships being denied, what same-sex relationships *are* is itself fundamentally not understood by some.

Consequently, the Shared Reasons Constraint seems overexclusive even if we endorse its weak variant. Recent social epistemology indicates that when members of privileged groups are profoundly ignorant of vulnerable groups' daily experiences, they are liable not to understand important considerations raised by members of such groups. Often, as the phenomenon of hermeneutical injustice suggests, this is because they do not grasp crucial concepts that figure in these considerations. When this happens, even the weak Shared Reasons Constraint prohibits appealing to such considerations. But some of these excluded reasons, including those discussed above, intuitively should have a place in public discourse. Excluding them silences marginalised groups, and thereby prevents them from contesting an unjust status quo. More precisely: excluding the important reasons in question undercuts public contestation's efficacy at advancing justice with respect to the policy issue on which the reasons were taken to bear. Therefore, there are powerful reasons to *resist* the Shared Reasons Constraint grounded in the value of freedom as non-domination.

This result is unfortunate from the standpoint of freedom: pro tanto reasons grounded in non-domination support the Shared Reasons Constraint in some respects, and pull against it in others. On the one hand, receiving justifications in terms of non-

³⁰ E.g., Rothman (2015).

³¹ E.g., Walker (2014).

shared reasons reduces individuals' ability to contest exercises of power (2.2). On the other, requiring that individuals appeal to shared reasons holds the political agency of vulnerable participants hostage to the ignorance or ill-will of others (2.3). Either way, it seems we fail to satisfy important freedom-based reasons.

In what follows, I critically assess the three most popular responses that deliberative democrats have developed to offset the problem of overexclusiveness. While these strategies are commonly presented together, they are conceptually distinct. Consequently, I will articulate them separately, to better discern the difficulties associated with each. I will, however, also evaluate how they might work jointly to overcome certain difficulties. In evaluating them, the question to bear in mind is this: how can we avoid making public deliberation overly exclusive, while responding as far as possible to the initial reasons for adopting the Shared Reasons Constraint? How, put differently, can we best weaken the trade-off between competing reasons grounded in the value of non-domination?

3. The Limits of Three Influential Responses

3.1. Deriving new reasons from old ones

An initial response acknowledges that the Shared Reasons Constraint may initially exclude important reasons, but insists that these considerations can become shared, and therefore admissible, via deliberative processes. Hence, the objection that some important reasons are *currently* excluded is not decisive against the Shared Reasons Constraint.

Krause describes the process through which new considerations are to be introduced as follows:

Public reason [...] always operates in medias res. [...] Sometimes new principles are introduced into public dialogue, principles that previously were external to public reason. Yet in order for a new principle [to be introduced] it must connect up with some principles that are already subject to public agreement³²

³² Krause (2008, 160-161). See also Rawls (1997, 784-786), Habermas (2008, 130-131), and Morgan-Olsen (2010, 235-238).

The idea, it seems, is that we introduce new reasons by showing that already-shared reasons support or make sense of these novel considerations. The process has a broadly inferential structure: new considerations are derived from, or translated in terms of, reasons that are already shared. Thus, through logical argument, public debate can produce a more adequately inclusive set of publicly-admissible reasons.

This derivation process notably echoes Rawls's famous proviso in his "wide" understanding of public political culture. According to Rawls's proviso, citizens may appeal to reasons that are not yet shared or public provided that, in due course, they give properly public reasons to support those appeals.³³ A religious citizen could permissibly appeal to the Gospel story of the Good Samaritan, Rawls suggests, so long as they "do not stop there, but go on to give a public justification for this parable's conclusions in terms of political values".³⁴ The recurring thought is that justifications in terms of non-shared reasons can be brought into the public sphere on the condition that they are shown to link up adequately to already-shared considerations.

While this proposal—to derive new reasons from already-shared ones—is *prima facie* appealing, it is unclear how fruitful it would be in settings marked by deep divisions and injustices. The proposal envisions the democratic process as involving public debate that systematically improves itself. We start from an overexclusive set of publicly-admissible reasons, and thence derive novel reasons, which in turn enable us to integrate further legitimate concerns into public debate. But the expectation that public debate will be self-improving in this way is premature. If the initial set of shared reasons is sufficiently small and biased against certain participants—perhaps because others are ignorant of their experiences—the derivation process could instead be viciously circular, in the sense of reproducing the initial exclusiveness of public deliberation.³⁵ To see this, imagine a highly inegalitarian society where the majority believes in a natural hierarchy based on sexual preference. Because equality and related moral concepts are not shared in this scenario, logically connecting the consideration that same-sex couples

³³ Rawls (1997, 784-785).

³⁴ *Ibid.*, 786. Note that Rawls's proviso is slightly ambiguous. The shared reasons introduced in due course could be introduced *either* to support the non-shared considerations (as I have been suggesting) *or* directly to support the policy that the non-shared reasons aimed to justify. However, the latter option seems unsatisfactory: it allows deliberators to appeal to non-shared considerations that are dear to them *only when* these are redundant from a justificatory perspective. That is, it suggests that non-shared reasons can be given to support P only when some other consideration already sufficiently supports P. But citizens who are concerned that morally-important considerations are excluded from public discourse do not simply want to *utter* those considerations in public. They want to use those considerations to *justify* policies.

³⁵ Bohman (2003, 770).

should have the right to marry to existing shared reasons would be extremely difficult, if not impossible. Here, the initial exclusiveness of public discourse would prevent some valid concerns from being derived from already-shared reasons.

One might reply that I am being too pessimistic. Even if *in principle* the derivation strategy can be inadequate when starting from highly exclusionary conditions, one might think that actual conditions are not like this. For instance, Krause observes that even in pluralistic polities like the United States, many evaluative concerns *are* shared. Indeed, equality, liberty, and the pursuit of happiness are enshrined in the US Constitution.³⁶ This, it might seem, makes it possible logically to derive many new reasons through public deliberation. ‘If we truly care about liberty and equality,’ proponents of same-sex marriage can argue, ‘then we are logically committed to respecting different lifestyles and recognising an equal right to marriage for same-sex couples.’

However, this reply overlooks another obstacle that stands in the way of deriving novel reasons. Despite invoking liberty and equality, the US Constitution was long interpreted as legitimating slavery. Likewise, in *Plessy v. Ferguson*, the US Supreme Court upheld the constitutionality of racial segregation by propounding the “separate but equal” doctrine. These cases remind us that even when a consideration x is widely recognised, this may well conceal deep disagreement over its correct interpretation. So, deriving a novel reason in a way that is likely to make it become shared requires not just connecting it with x , but more specifically connecting it to the audience’s interpretation of x . What is worse, *Plessy v. Ferguson* highlights that widely-held prejudices—say, beliefs in a natural hierarchy—may infect ‘promising’ values, like equality, so that they come to be interpreted in a manner consistent with existing injustices. Thus, even when a reason seems shared, disagreement over its interpretation and exclusionary interpretations may combine to obstruct attempts at analysing legitimate concerns in its terms.

According to social epistemologists concerned with the social production of ignorance, this last difficulty is a common feature of non-ideal, divided societies.³⁷ Mills, for example, suggests as much when commenting on contemporary understandings of equality and racism:

³⁶ Krause (2008, 157, 160).

³⁷ E.g., Mills (2007), Fricker (2013), Stanley (2015, ch. 3), Talisse (2009, 63-64).

If previously whites were color demarcated as biologically and/or culturally unequal and superior, now through a strategic ‘color blindness’ they are assimilated as putative equals to the status and situation of nonwhites on terms that negate the need for measures to repair inequities.³⁸

Mills is saying that although equality is now widely affirmed, many understand it in colour-blind terms. Roughly, on this conception, racial equality is realised when no one is treated differently because of their race. This interpretation, Mills suggests, is incompatible with many concerns that black activists have defended on grounds of equality. To those who endorse a colour-blind conception of equality, the consideration *that blacks are owed differential treatment (including reparations for historical injustices)* cannot intelligibly be derived from considerations of equality. Indeed, as Mills notes, to them the real opponents of equality and “the real racists are the blacks who continue to insist on the importance of race”.³⁹ Thus, such individuals will reject attempts at translating demands for reparations in terms of equality.

For another example, consider again same-sex marriage. Suppose an advocate of same-sex marriage wanted to introduce the non-shared consideration *that same-sex relationships are valuable* into public discourse. And suppose she did so by connecting same-sex relationships to the idea of a loving relationship, where her opponents recognised *that loving relationships are valuable*. In response to similar arguments, opponents of same-sex marriage sometimes respond that true romantic love is essentially a relation between a woman and a man. Some religious believers maintain this restrictive concept of love because, citing Corinthians, they claim that true love is essentially godly, while same-sex relationships are sinful.⁴⁰ Alternatively, if one believes that loving relationships are inherently healthy and valuable, but (in line with the DSM’s pre-1973 classification) that homosexuality is a disorder, then one is bound to think that same-sex relationships cannot be subsumed under the concept of a loving relationship. Still others might object, as some philosophers notoriously have, that same-sex relationships, in virtue of instantiating a form of narcissistic desire, are perversions of genuine love.⁴¹ More generally, the problem is the following: when deliberators hold

³⁸ Mills (2007, 27-28).

³⁹ Ibid., 28.

⁴⁰ Halley (2017).

⁴¹ Scruton (2001, 307-310). See, similarly, Finnis (2008, 390-396).

such contrasting concepts of love, there is no shared understanding of ‘loving relationship’ from which one might derive the consideration that same-sex relationships are valuable.

These considerations suggest that, in divided settings, deriving novel reasons from the existing pool of shared considerations is very challenging. Although there may seem to be agreement on some shared reasons, this agreement is often more apparent than real: it conceals profound disagreement on how these considerations should be interpreted; and it obscures the fact that, in unjust circumstances, even seemingly promising considerations may be interpreted in ways that accommodate injustices. When the initial pool of genuinely shared reasons is highly limited, so too are the prospects for deriving new considerations on their basis.

This is not to deny that novel reasons *do* over time enter public discourse and do become shared. Rather, the point is that if we want to conceptualise and prescribe processes through which they might do so, the strategy of logically deriving new reasons from old ones seems insufficient on its own. Thus, let us see how we might try to supplement this proposal.

3.2. Restricting the Constituency

When very few reasons are already genuinely shared, the prospects for deriving new shared reasons from old ones seem limited. In response, one solution aims to facilitate processes of derivation by adding some important considerations that not all actual citizens share to the set of publicly-admissible reasons. The idea is that, if the initial set of publicly-admissible reasons is enriched thus, it will be easier for participants to voice their demands and to introduce new important considerations on the basis of already-admissible reasons. Indeed, they will have more resources from which to derive novel considerations.

This strategy is most clearly exemplified by public reason theorists, particularly those working in the Rawlsian tradition, when they restrict the constituency of the Shared Reasons Constraint. The Shared Reasons Constraint, recall, says that deliberators must appeal to reasons that are shared by all participants when publicly justifying their claims. So far, I have been assuming that ‘participants’ refers to all *actual* adult citizens. But public reason theorists characteristically adopt a more restricted

understanding of the Shared Reasons Constraint's constituency, by taking 'participants' to refer to *idealised* persons. Participants can be idealised in two ways: epistemically, when they are better at reasoning and responding to evidence than actual participants; and morally, when they have better values and motivations than actual participants. Note that although the parties are idealised, some subset of actual participants may satisfy the idealising conditions, while others may not. Often the former are labelled "reasonable", and the latter "unreasonable".⁴²

Rawls influentially illustrates this approach.⁴³ He claims that the reasons given when justifying coercive power must be ones that "free and equal citizen[s] might *reasonably* accept".⁴⁴ To be reasonable involves, firstly, a moral idealisation. Reasonable participants view others as free and equal citizens and are therefore prepared to offer them fair terms of cooperation. Moreover, because reasonable citizens recognise the "burdens of judgment"—the many hazards that inevitably confront human reasoning—they do not expect or demand that others hold ethical views identical to their own.⁴⁵ According to Martha Nussbaum, there is evidence that Rawls's account of reasonableness also involves an epistemic idealisation: Rawlsian reasonableness, on her interpretation, requires upholding views that are basically coherent and responsive to new evidence.⁴⁶

That being said, many political theorists working outside of the Rawlsian tradition adopt a similar strategy. Krause, for instance, suggests that the test for whether a consideration is admissible in public deliberation is not whether it is endorsed by all actual citizens, but rather whether it could "be endorsed from within a suitably structured moral standpoint." What makes concerns "endorsable in this context is that they can be affirmed [...] from within a deliberative perspective that manifests equal respect."⁴⁷

⁴² Quong (2017, Section 3).

⁴³ Rawls (1997, 769-771). See also Quong (2017, Section 3).

⁴⁴ Rawls (1997, 770), emphasis added.

⁴⁵ Ibid.

⁴⁶ Rawls (1993, 58-59). See Nussbaum's (2011, 24-25) discussion.

⁴⁷ Krause (2008, 163; see also 126, 132, 166-167). Additionally, see Estlund (2008, 4) and Gutmann and Thompson (1996, 55-56). A terminological note: Rawls, Quong, and Estlund are often labelled 'public reason theorists', whereas Krause and Gutmann and Thompson are frequently described as 'deliberative democrats'. Typically, the latter care essentially about *actual* deliberative justification, whereas the former care primarily about the *justifiability* of institutions, though they often also care about actual justification. What matters here is that both restrict the Shared Reasons Constraint's constituency. And since restricting its constituency is commonly associated with public reason theorists, I will generally use 'public reason theorist' to refer to

Thus, though Rawls's account is the most influential, different theorists have offered many competing accounts of what the right level of idealisation is (and by extension, of what exactly constitutes reasonableness).⁴⁸ Consequently, the arguments that follow will not depend on the specific moral and epistemic conditions Rawls builds into his idealisation. Instead, the moral costs that I will argue are associated with idealising the participants who constitute the Shared Reasons Constraint's constituency should apply *to the extent that* we idealise them.

Restricting the Shared Reasons Constraint's constituency seems *prima facie* helpful when it comes to addressing the overexclusiveness problem. By excluding 'unreasonable' participants who are insensitive to important descriptive or normative considerations when interpreting the Shared Reasons Constraint, idealisation prevents such actual persons from ruling out appeals to these considerations. Put differently, this strategy prevents the content of public deliberation from being held hostage by the perspectives of unreasonable citizens.

However, restricting the constituency to reasonable persons is problematic in two respects. The first and most widely discussed worry concerns the fate of unreasonable people. Some theorists express the concern that excluding unreasonable persons from the Shared Reasons Constraint's constituency means stripping them of their basic political status: unreasonable persons, Marilyn Friedman suggests, "will be treated like the bearers of a pestilence [...] They will be denied the full protection of the system's basic rights and liberties, particularly freedom of expression."⁴⁹

As Quong has emphasised, however, Friedman's articulation of this initial worry rests on a misinterpretation. Excluding unreasonable persons from the Shared Reasons Constraint's constituency does not mean excluding them from the political community of citizens altogether. Actual unreasonable citizens, Quong insists, generally retain their basic legal and political rights, including the rights to vote, to stand for office, and even to publicly express unreasonable views.⁵⁰ Indeed, recall that the Shared Reasons Constraint is a moral norm, not a legal one. So, contra Friedman, it does not necessarily require that anyone be *legally* prevented from expressing their views.

advocates of this restriction. On the relation between justifiability and actual deliberative justification, see Vallier (2015).

⁴⁸ For discussion of this diversity, see Vallier (2011, 371-372) and Quong (2017, Section 3).

⁴⁹ Friedman (2000, 23).

⁵⁰ Quong (2011, ch. 10)

Unreasonable citizens, then, are not formally excluded from the democratic political community. Where they *are* discounted is in the theoretical process of determining what qualifies as a shared reason. What does this mean for actual non-ideal deliberative politics? Because they are discounted, some reasons that are not actually shared by unreasonable citizens will nevertheless count as shared. Citizens who embrace these not-actually-shared reasons will therefore be welcome, morally speaking, to appeal to them when justifying their political demands. Hence, to the extent that the Shared Reasons Constraint is realised, public deliberation will partly be framed in terms that unreasonable citizens do not accept and eventually do not understand. Moreover, although the Shared Reasons Constraint is not a legal norm, we saw in 2.1 that the moral reasons underpinning it may warrant lighter, non-coercive ways of upholding it. This may involve, say, ignoring the public complaints of unreasonable citizens, should they object that they cannot understand or accept some reasons that are being publicly appealed to. Indeed, such complaints are irrelevant if unreasonable citizens are not part of the constituency for whom considerations must be shared.

In this light, even if Quong is right that restricting the constituency does not strip unreasonable citizens of formal citizenship rights, doing so nonetheless imposes a moral cost on them. Section 2.2 argued that the Shared Reasons Constraint is pro tanto desirable because it advances the freedom as non-domination of individuals subjected to power. Requiring that coercive power be justified using shared reasons enhances subjects' capacity to contest that power. But consider now the position of unreasonable parties. Some, perhaps many, of the reasons raised in deliberative justifications are now unacceptable and eventually unintelligible to them. Consequently, they are relatively disempowered with respect to these justifications. Moreover, their objection to the presence of these reasons may eventually be ignored as irrelevant. Hence, it seems the Shared Reasons Constraint no longer protects *them* from domination. Insofar as this is so, idealising the Shared Reasons Constraint's constituency undercuts the original rationale for that constraint.⁵¹

Now, this initial problem does not entirely preclude idealising the Shared Reasons Constraint's constituency. As 2.3 discussed, the overexclusiveness resulting from the non-idealised Shared Reasons Constraint in conjunction with the ignorance or immorality of some citizens (the unreasonable) exposes other citizens to domination

⁵¹ For a similar point, see Enoch (2015).

(those whose legitimate concerns are misunderstood or misrecognised). It may well be that sometimes, protecting the latter from domination warrants, all things considered, subjecting the former to domination. This is the conclusion Matt Sleat comes to. While he emphasises that the unreasonable are dominated when public justification is geared towards reasonable participants, his conclusion is not that this is impermissible. Instead, it is that violations of freedom are an unavoidable, if regrettable, feature of non-ideal politics.⁵²

This first concern, then, does not imply that restricting the constituency necessarily generates *insuperable* moral costs.⁵³ Nevertheless, it does show that this proposal is problematic for our purposes: our aim, as stated when concluding Section 2, is to *weaken* the trade-off between dominating those whose legitimate concerns are not publicly recognised, and dominating those who do not recognise those concerns. But restricting the constituency does nothing to weaken this trade-off. Instead, it suggests how, *given* this trade-off, we can minimise domination. Accordingly, it should be a last resort, introduced only once we have exhausted alternative solutions aimed at weakening the trade-off.

Idealising the Shared Reasons Constraint's constituency generates a second, more serious, problem relating to non-ideal conditions. Existing democracies are non-ideal not just in that some individuals have unreasonable views, but also in that there is often a mismatch between who is right and who has power. Often, the unreasonable are also powerful. This, recall, is a recurrent theme of contemporary social epistemology: as the examples of marital rape, homophobia, sexual harassment, and micro-aggressions indicated, it is often the experiences and concerns of *vulnerable* groups that are misrecognised and it is *dominant* groups who misrecognise them. Public reason theorists sometimes obscure this point by focusing on cases where the unreasonable are marginal—for instance, religious extremists like the Westboro Baptist Church.⁵⁴

If the unreasonable are also powerful, however, idealising the Shared Reasons Constraint's constituency is problematic. Such a norm welcomes debates that are framed in terms that unreasonable citizens may neither accept nor understand, and

⁵² Sleat (2013, 362-363).

⁵³ One might think that domination is not morally costly at all when it affects unreasonable citizens. I follow Waldron (2012, 146-150), Enoch (2015, 123), and Raz (1998, 32-34) in resisting this view. Indeed, *ceteris paribus*, a world where we can protect vulnerable groups without dominating the unreasonable—say, by convincing them instead—seems morally better.

⁵⁴ E.g., Blake (2005, 233), Gutmann and Thompson (1996, 63-69).

recommends ignoring their complaints concerning this framing. But this seems ill-advised if they are dominant. If we fail to engage with the objections or concerns of powerful agents, we are unlikely to achieve political change. As mentioned earlier, King's discourse was effective partly because, rather than ignore the concerns of the (often bigoted) white majority, he tried to speak a language they were responsive to. Similarly, Abraham Lincoln often appealed to legal grounds for ending slavery precisely because many of his Congressional interlocutors were unresponsive to the moral case.

Hence, restricting the Shared Reasons Constraint's constituency jeopardises one of the prime reasons for advocating inclusive deliberation. Such processes help hold the powerful accountable. By demanding justifications for their decisions, and publicly scrutinising those justifications, weaker parties can publicly shame the powerful and thereby exercise some control over them. But if weaker agents frame their claims in terms of considerations that powerful agents are unresponsive to, or simply ignore powerful agents, public deliberation cannot serve this purpose. Hence, it is not just the non-domination of unreasonable citizens that is problematic for the present proposal. It is also the non-domination of vulnerable groups.

Some public reason theorists might reply that not addressing this second worry is not an oversight. Quite simply, they are interested in theorising public reason for ideal rather than non-ideal conditions. Quong, like Estlund and Rawls, explicitly suggests this. He remarks that

the idea of public reason and the duty of civility belong to an *ideal* of how things ought to be in a democratic society, assuming citizens as just and society as well-ordered [...] Pointing out that adhering to the duty of civility may be counterproductive under current conditions is thus no objection to the ideal, since current democratic societies are far from ideal, or well-ordered.⁵⁵

There are two things to say here. Firstly, although this is a possible response, some political theorists who idealise the Shared Reasons Constraint's constituency clearly are interested in non-ideal theory. For example, Gutmann and Thompson as well as Krause

⁵⁵ Quong (2013, 270; 2017, Section 8.1). See also Estlund (2008, 19) and Rawls (1997, 766n3). As Enoch (2015, 125n37) observes, however, Quong (2011, 158-159) occasionally speaks as if public reason norms applied in non-ideal conditions.

intend their discussion of public discourse to be relevant to actual conditions. Accordingly, they frequently cite actual cases—for instance, pronouncements by the Civil Rights Movement—to illustrate why they defend certain discursive norms.⁵⁶

Secondly, while this response might shield Rawlsian public theorists and other ideal theorists from my criticism, it does so only by making their topic of inquiry distinct from my own. Hence, it does not undermine my point that, in non-ideal theory, restricting the Shared Reasons Constraint's constituency generates significant problems. Even if those authors were not concerned with non-ideal theory, it remained an open question (which Quong acknowledges to be “hugely important” and underexplored⁵⁷) whether their recommendation could also be prescriptive there. I have reached the conclusion that in important respects, it is not.

In summary, the proposal that we restrict the Shared Reasons Constraint's constituency to idealised or ‘reasonable’ participants faces two serious limitations (which echo the limitations of Chapter 1's Moral Imposition Solution) when it is applied to public deliberation taking place in non-ideal conditions. Firstly, the proposal is insensitive to non-ideal contexts of power: when unreasonable parties are also powerful, the idealised Shared Reasons Constraint jeopardises the freedom as non-domination of vulnerable groups. Where the unreasonable are *not* powerful, the proposal might seem more viable, as they can safely be ignored. However, even in such conditions restricting the Shared Reasons Constraint's constituency seems problematic: it risks subjecting unreasonable citizens to domination. While this might sometimes be a necessary cost to protect misrecognised vulnerable groups from domination, we should only accept this solution as a last resort. This is because it does not weaken the trade-off between the non-domination of the unreasonable and of the reasonable. Rather, it simply suggests what the best decision is *given* that the trade-off exists. Prior to idealising the Shared Reasons Constraint's constituency, then, we must investigate whether there are ways of weakening or dissolving this trade-off.

⁵⁶ E.g., Gutmann and Thompson (1996, 134), Krause (2008, chs. 4-5), Dreben (2003, 317, 328).

⁵⁷ Quong (2013, 270).

3.3. Adopting a Systemic View of Deliberative Sites

According to the third popular solution, the overexclusiveness problem seems worrying largely because I fail to distinguish between different sites of the deliberative ‘system’. On the systemic understanding of deliberation, we should not imagine democratic deliberation as involving a single deliberative forum, which aspires to include all citizens. Instead, we should understand deliberative democracy as a whole *system* composed of smaller parts—i.e., composed of many arenas or sites for public discussion, which vary in size, function, and formality, and which are connected more or less directly to one another.⁵⁸

Systemic approaches typically distinguish two main kinds of deliberative sites. Formal public arenas, such as legislative bodies, involve deliberation which leads to coercively-enforced decisions. By contrast, in more informal arenas, often situated in civil society, deliberation largely helps participants form considered opinions on political issues.⁵⁹

There are strong reasons for endorsing the systemic view of deliberative democracy. Firstly, interpreting deliberation as a system composed of many interconnected sites helps us “think about deliberative democracy in large-scale societal terms”.⁶⁰ Accordingly, it contributes to addressing the criticism that deliberative democracy’s emphasis on inclusive deliberation is unrealistic when applied to large-scale mass democracies. Secondly, by describing democracy as comprising interconnected arenas which differ in size, formality, and function, it seems capable of accommodating naturally the intuition that institutions like consciousness-raising groups, partisan media outlets, and protests can enhance the quality of democracies, even though individually, such institutions are not fully welcoming to all citizens. Because of this twofold appeal, we should embrace the systemic approach to deliberative democracy. Nevertheless, I aim to challenge the more specific claim that the systemic approach helps solve the overexclusiveness problem.

Why think that it helps? The systemic view of deliberation states that different deliberative sites vary in formality, so that some are regulated by more constraining

⁵⁸ Mansbridge et al (2012).

⁵⁹ E.g., Habermas (1996, 307-308), Krause (2008, 121).

⁶⁰ Mansbridge et al (2012, 2).

discursive norms than others.⁶¹ In this light, a seemingly straightforward strategy for offsetting the Shared Reasons Constraint's exclusiveness recommends confining it to formal deliberative sites. Rawls, for instance, affirms that "the idea of public reason does not apply to the background culture [i.e., civil society], with its many forms of non-public reason nor to media of any kind". It only applies to what he calls the "public political forum", which includes the discourse of judges, government officials, legislators, and candidates for public office.⁶² Krause is even clearer in suggesting that restricting the site of the Shared Reasons Constraint should mitigate the problem of overexclusiveness. While "laws and policies that are enforced on all of us need to be capable of being justified in terms of principles and concerns that are shared [,] [t]he mistake that is often made here is to treat all deliberation, including opinion-formation", in this way.⁶³ It is entirely permissible, Krause suggests, for citizens to appeal to non-shared reasons in informal discourse aimed at opinion-formation. Thus, by distinguishing between formal and informal sites of decision-making, "a deliberative-system approach [...] will be less vulnerable to the charge, frequently levied against deliberative democrats as well as liberals that regulatory ideals of public deliberation systematically exclude the deeply held beliefs and values of many citizens".⁶⁴

On this view, then, there is a division of labour within the deliberative system, which weakens the trade-off articulated at the end of Section 2. The freedom as non-domination of vulnerable and misunderstood groups is preserved by waiving the Shared Reasons Constraint in informal opinion-formation settings, so that their perspectives are included there. Concurrently, the pro tanto freedom-related benefits of having the Shared Reasons Constraint are maintained by applying it to formal decision-making arenas.

The devil is in the details. How is this division of labour supposed to work? An initial proposal is that simply because the system has a part—informal public spheres—that includes many viewpoints and reasons, including non-shared ones, we can consider the system as a whole to be sufficiently inclusive. This is unsatisfying because those who press the complaint of overexclusiveness do not merely want to express their important concerns in informal public fora. They want the expression of those

⁶¹ Dryzek (2010, 7).

⁶² Rawls (1997, 768).

⁶³ Krause (2008, 156).

⁶⁴ Ibid., 121.

concerns to influence formal decision-making that produces coercive decisions. Only if this condition is satisfied will the inclusion of their concerns allow them to contest policies and to protect themselves from domination.

So, a process must exist whereby the novel considerations and arguments developed in the informal sphere are transmitted to formal decision-making spheres. But since the formal decision-making sphere *does* have a Shared Reasons Constraint, this transmission would require translating those novel considerations in terms of reasons that are shared! The problem, then, is that the solution under consideration seems simply to reproduce the difficulties examined in 3.1. Confining the Shared Reasons Constraint to formal spheres is only attractive if the novel reasons raised in informal sites can make their way into formal deliberative sites. And this, in turn, requires deriving those reasons from already-shared reasons. Since we have been seeking ways to supplement or enhance the idea of deriving new reasons from old ones, this seems unhelpful.

Here, Krause advances a more promising suggestion. She suggests a sense in which distinguishing between different sites in the deliberative system might facilitate the derivation process, rather than simply reproduce its difficulties:

Through the layered process of public debate in the many different venues of the deliberative system, the arguments defending particular policies will either gain wider acceptance, or they will fade from the public agenda. By the time a policy is ready to be voted up or down we can expect—and require—that the arguments [...] do appeal to sentiments and concerns that are generally shared⁶⁵

The idea, as I understand it, is that unconstrained public deliberation in the informal sphere can allow non-shared reasons to become shared *prior* to an issue being deliberated in the formal sphere. This seems to facilitate the process of deriving new reasons from already-shared ones: if prior informal discussion has helped popularise the concerns of marginalised groups, it will presumably be easier to derive their claims from already-shared reasons when trying to introduce them into formal decision-making arenas.

⁶⁵ Ibid., 121-122.

Ultimately, however, even this more promising suggestion does not avoid the problem at hand. It explains why the division of labour between different deliberative sites facilitates the transmission of new reasons from the informal to the formal public sphere—namely, that deliberation in the informal sphere augments the stock of shared reasons which can then be used to derive new considerations in the formal sphere. But in doing so, it generates a new and isomorphic ‘derivation’ problem. We can now ask how discourse *within* the informal public sphere enables non-shared reasons to become shared by those who initially rejected them. If those wishing to popularise their concerns simply appeal to non-shared considerations (as the informal setting allows them to do), then it is unclear, absent further considerations, why those who do not already recognise these considerations would come to do so. Indeed, arguing for same-sex marriage by appealing to the non-shared consideration that same-sex relationships are valuable loving relationships will presumably leave unmoved those who reject this consideration. If, however, they try to logically derive these non-shared reasons using already-shared reasons, then we are back where we started.

As such, appealing to the division of labour between different deliberative forums is unsatisfactory in the present context. This is not to disavow the systemic conception of deliberative democracy itself. As I stressed earlier, we should embrace the systemic view of deliberation, along with its intuitive proposal that the Shared Reasons Constraint should only apply to more formal deliberative arenas. This view will prove important in subsequent chapters. My point here is more limited: it is unclear, absent further considerations, how adopting this understanding of deliberation helps address the overexclusiveness problem.

I have considered the three most influential strategies for offsetting the overexclusiveness of the Shared Reasons Constraint, and have suggested that they encounter significant limitations, both individually and considered together. While logically deriving non-shared reasons from already-shared reasons identifies a plausible way of introducing novel reasons into public deliberation, its prospects seem limited if we start from non-ideal contexts where—as social epistemologists have suggested—some groups lack the experiences and conceptual resources needed to appreciate others’ valid concerns (3.1). To bolster this strategy, one might restrict the Shared Reasons Constraint’s constituency to idealised participants. Indeed, doing so expands the pool of initial shared reasons from which the derivation of new reasons then

proceeds. This proposal, however, is poorly equipped to handle non-ideal circumstances where those who are insensitive to legitimate concerns are also empowered (3.2). Finally, confining the Shared Reasons Constraint to formal sites in the deliberative system only postpones the problems that the ‘derivation’ strategy faces (3.3).

4. Expanding the Form of Democratic Public Deliberation

To alleviate these difficulties, I will now propose a further strategy for offsetting the Shared Reasons Constraint’s overexclusiveness. I want to suggest that this overexclusiveness concerning what *contents* are admissible in public discourse is partly sustained by an overly restrictive conception of what *forms* of communication—i.e., what vehicles for communicating propositional contents—are appropriate. To remedy this, non-argumentative and emotionally-charged forms of speech should occupy a more central role in our normative account of formal and informal democratic public deliberation. Note that my primary aim here is to sketch how such a discursive norm, which expands the forms of speech that can be used publicly to express propositional contents, might complement the strategies canvassed in Section 3 and weaken the Shared Reasons Constraint’s overexclusiveness. However, I will not develop this solution at great length here. Instead, I will develop it in Chapter 4, where I specifically discuss the epistemic merits of angry public narratives.

4.1. Beyond Argument

The three strategies considered in Section 3 suggest a picture of democratic public discourse characterised by broadly *inferential* forms of discourse, like argument, where deliberators derive conclusions from premises via logical steps. This is explicit in the idea of deriving new considerations from already-shared ones (3.1). Furthermore, the strategy of restricting the Shared Reasons Constraint’s constituency (3.2) also works within this picture: it is a way of expanding the set of contents from which the derivation of new shared reasons begins. So, it does not challenge the notion that speakers must logically connect novel considerations to already-shared considerations. As for confining the Shared Reasons Constraint to formal deliberative sites, we have

seen that this strategy ultimately reintroduces the need to derive new reasons from already-shared reasons (3.3).

This picture of democratic discourse is incomplete. In addition to argument, democratic public deliberation should welcome forms of discourse, such as narrative, that are non-argumentative and typically charged with emotion. *Narrative* is discourse that aims to transmit one's experience or perspective on the world. This can consist of personal testimony—i.e., telling others about one's experiences of an unusual event or practice. But it also includes more general perspective-sharing—i.e., telling a story to get people to see a common event or practice as one does.⁶⁶ As I define it, *emotionally-charged* narrative can involve not just positive emotions such as joy, but also more contentious emotions, such as anger and despair.

Importantly, these forms of discourse are not *inferential*: they do not involve logically connecting what is said to premises that are currently shared. This is promising in light of the difficulties encountered earlier. What conclusion one can reach via logical inference depends fundamentally on what premises one starts from. This is particularly clear in the case of deductive argument, which is non-ampliative: there can be nothing in the conclusion that is not in some sense already in the premises. That is why, as we saw in 3.1, the strategy of deriving new reasons from already-shared reasons seems unpromising when it starts from an extremely restricted set of shared considerations. In contrast, since emotionally-charged narrative is not inferential, it is not hostage to the lack of shared premises. It aims directly to introduce or make visible certain experiences or perspectives, and thereby to enrich the pool of publicly-recognised considerations.

The next section will further outline how the epistemic value of emotionally-charged narratives helps address the overexclusiveness problem. Before proceeding, we should consider how the present recommendation fits into broader debates in democratic theory. The idea that we should give a greater role to emotionally-charged and non-argumentative forms of speech, such as narrative, is by no means new. It has long been championed by democratic theorists concerned with non-ideal theory, most notably Iris Marion Young. But these theorists, as mentioned in the introduction, are often sceptical of discursive constraints like the Shared Reasons Constraint.⁶⁷ By

⁶⁶ Sanders (1997, 369-373); Young (2000, 70-77).

⁶⁷ Young (1996, 131-132). See also Sanders (1997, 372), Mouffe (1999, 755), and Dryzek (2000, 73-74). Dryzek (2010) is more open to discursive constraints than some of these other theorists. But he nonetheless claims that “anything goes [...] provided that it is (a) capable of inducing reflection, (b) noncoercive, and (c)

contrast, I am arguing that expansive norms relating to the form of democratic speech are needed to *complement* the Shared Reasons Constraint, by offsetting its overexclusiveness.

On the other hand, deliberative theorists who *do* support the Shared Reasons Constraint have traditionally been less accommodating towards emotionally-charged and non-argumentative forms of speech. In fact, Young's defence of such discursive forms is a critical response to classical theories of deliberative democracy. Traditional deliberative democrats, Young suggests, aim to stave off the Platonic objection that democracy hands power over to the mob's unruly passions. To do so, they adopt 'rationalist' norms of deliberation: they affirm that proper public communication consists not of emotional outbursts or intoxicating rhetoric, but of calm, logical arguments from common premises.⁶⁸ Unlike the Shared Reasons Constraint, which governs the content of public reason-giving, Young emphasises that these exclusionary norms govern its *form*. They exclude people's contributions "from serious consideration not because of what is said, but how it is said".⁶⁹

There is certainly some evidence supporting Young's charge among the most influential defenders of the Shared Reasons Constraint, most prominently (though not exclusively⁷⁰) Rawls and Habermas. When discussing the idea of public reason, for instance, Rawls distinguishes between "reasoning"—which is governed by "principles of inference"—and "rhetoric or means of persuasion"—which are not. Having done so, he immediately stresses that he is concerned with former.⁷¹ A similar observation is regularly made concerning Habermas. Habermas famously defends deliberative democracy as a way of subjecting power to the "unforced force of better argument".⁷² And he interprets emotions as essentially non-cognitive states, whose political expression seems relatively lacking in epistemic value.⁷³ On the basis of such

capable of connecting the particular experience of an individual, group, or category to some more general principle" (34). This is weaker than the requirement that deliberators appeal to shared considerations, as some general principles may not be shared (*ibid.*).

⁶⁸ Young (2000, chs. 1-2). For further critical discussion of these rationalist norms, see Sanders (1997) and Garsten (2011, 162-167).

⁶⁹ Young (2000, 56).

⁷⁰ For other classical deliberative democrats who endorse rationalist norms, see Cohen (2009, 248-251), Elster (1998, 109), and Spragens (1990, 128). Benhabib (1996, 82-83) recognises that affective storytelling has a place in the deliberative ideal, but restricts it to informal arenas.

⁷¹ Rawls (1993, 220).

⁷² Habermas (1996, 306).

⁷³ *Ibid.*, 294.

comments, some theorists conclude that Rawls and Habermas openly exclude emotionally-charged and non-argumentative discourse from their deliberative ideals.⁷⁴

Still, one might harbour reservations. Although these comments about emotions and rhetoric are suggestive, they tend to be made in passing and are not developed at length. So, one might worry that it is uncharitable to suggest that Rawls and Habermas, and the deliberative theorists following in their wake, *explicitly exclude* emotionally-charged narrative from their normative ideals.

Instead, as Dryzek suggests, a fairer charge is one of silence: even if they do not explicitly exclude emotionally-charged and non-argumentative discourse, Rawls and Habermas have little to say about these forms of discourse in their normative accounts of deliberation.⁷⁵ Importantly, this second charge also applies to more recent defenders of the Shared Reasons Constraint, particularly Rawlsian public reason theorists. Quong, for example, almost never mentions emotional or narrative forms of discourse. The main exception to this is even more telling. While listing objections to the idea of public reason, he cites Young's concern that public reason "marginalizes [...] emotive, passionate, or rhetorical forms of discourse". But he promptly brackets this objection, and never returns to it.⁷⁶

On this evidence, the deliberative theorists who have most closely theorised the Shared Reasons Constraint commonly do not explicitly assign a central role to non-argumentative and emotionally-charged speech within their normative models of public deliberation. By contrast, my suggestion is that a commitment to the Shared Reasons Constraint should *motivate* such an expanded conception of deliberative forms. As the next section will discuss, emotionally-charged narrative performs a crucial epistemic function which helps offset the Shared Reasons Constraint's exclusiveness.

However, it is important to emphasise that there are significant exceptions to this tendency among defenders of the Shared Reasons Constraint. Some defenders of the Shared Reasons Constraint *do* recognise the importance of emotionally-charged narrative. Most notably, Krause explicitly combines the Shared Reasons Constraint with a firm endorsement of emotionally-charged narrative in formal and informal deliberation. And, unlike some other deliberative democrats who combine these two

⁷⁴ E.g., O'Neill (2002), Abizadeh (2007).

⁷⁵ Dryzek (2010, 69-70).

⁷⁶ Quong (2011, 260).

features,⁷⁷ she does this partly out of recognition of its epistemic value.⁷⁸ Hence, to clarify my argument, I will later return to the residual differences between her account and my own.

4.2. Emotionally-Charged Narrative and the Overexclusiveness Objection

How does emotionally-laden narrative help offset the Shared Reasons Constraint's overexclusiveness? By contrast with, say, deductive inferences from shared resources, narrative is an ampliative form of communication. By telling a story that gets people to experience the world in a certain way, it strives directly or non-inferentially to introduce new experiences and concepts into the public sphere.⁷⁹ For instance, in his *Invisible Man*, Ralph Ellison vividly depicts the perspective of a black American man who is persistently overlooked, misunderstood, slighted. Thus, he effectively reproduces those experiences, such that readers from other social groups may imaginatively perceive how daily slights cumulatively produce far more than a trivial offense. In this way, readers may come to see the irreducible feature of the world that the concept 'micro-aggression' aims to capture. So, emotionally-laden narrative helps *show* something when, because of a lack of shared experiences or concerns, it cannot yet convincingly be logically *argued for*.

What role does emotion play in this? Narrative could conceivably be expressed dispassionately. So why am I defending emotionally-charged narrative? Narrative seeks to increase the pool of shared experiences. But short of making people experience the

⁷⁷ E.g., Gutmann and Thompson (2004), Mansbridge (1999), Chambers (2004). Gutmann and Thompson (2004, 50-51) combine the Shared Reasons Constraint with acknowledgement of emotional speech's importance, but—as Young (2000, 67n11) and Krause (2008, 152-153) observe—they stress emotional speech's motivational rather than epistemic value, and classify it as a nondeliberative supplement to deliberation. Mansbridge (1999) also affirms that emotions play a positive role in deliberation (213) and expresses openness to Gutmann and Thompson's "reciprocity" requirement which requires appealing to mutually acceptable considerations (222). But the function of emotions she most emphasises again seems to be motivational: emotions matter because "reason can proceed only rarely without emotional commitment, *if only an emotional commitment to the process of reasoning*" (225, emphasis added). Similarly, although Chambers (2004) is sympathetic to the requirement that deliberators appeal to shared reasons (390-382), and although she asserts that emotions have an important role to play in public discourse (402-403), she predominantly describes emotional language as a "mobilizing tool" (402).

⁷⁸ See Krause (2008, 113-122, 162-165) on the epistemic value of emotionally-charged narrative. Note that another reason Krause deems emotion central to deliberation is less epistemic and more normative—on her view, reasons for doing things are constituted by affective concerns (2008, 150-151). Given this metaethical position, the process of deliberating about reasons for and against policies necessarily involves appealing to what we care about, affectively. For a similar point, see Hall (2007, 90-92).

⁷⁹ For evidence corroborating narrative's ability to publicise new experiences, see Polletta and Lee (2006, 718-719), Wesolowska (2007, 674, 676), Huijer (2009, 237).

actual relevant events themselves, this can only be done by getting them to imagine those experiences. However, even if someone describes an event accurately, imaginatively reconstructing it may be difficult. This is where emotion becomes epistemically significant. There is a widespread and involuntary tendency among human beings to resonate with or ‘catch’ the emotions of others. As Hume famously puts it: “as in strings equally wound up, the motion of one communicates itself to the rest, so all the affections readily pass from one person to another”.⁸⁰

For this reason, weaving emotion into an enumeration of facts plays an important epistemic role, as it facilitates the communication of experiences. By transferring emotion to one’s hearers, it helps them imagine those facts more vividly, in a way coloured by emotion. Chapter 4 will elaborate at length on the precise way in which the ‘colouring’ offered by emotions is epistemically valuable. For now, I leave things at the intuitive idea that emotions have a distinctive felt experience which highlights or renders more salient significant properties of our perceptions and imaginings. Thus, emotionally-laden narrative seems an effective means of getting the public to share one’s experience of an event or practice and of thereby enriching the conceptual categories people have for interpreting events and practices.⁸¹

Consider two examples relating to same-sex marriage, both drawn from French public debates in 2013. The first illustrates how emotionally-laden narrative in informal settings can communicate experiences and thereby enrich publicly-shared conceptual resources. Recall the deliberative impasse discussed in 3.1. In this example, defenders of same-sex marriage want to introduce the consideration that same-sex relationships are valuable by connecting such relationships to the idea of a loving relationship, but the concept of a loving relationship is itself disputed. Consider now *La Vie d’Adèle*, a film which gained wide attention against the background of public debates concerning same-sex marriage. The film portrays the relationship of two women, Adèle and Emma. Given our interest in how practices and claims are understood in public, observing the language critics used to characterise the film is instructive. *La Vie d’Adèle* was notably described by critics as “evok[ing] love in its purest and most passionate form”,⁸² as “brilliantly demonstrating how love feels”,⁸³ and finally as a “devastating portrayal of

⁸⁰ Hume (2009[1738], 3.3.1.7). See also Coplan (2011) and Chapter 4, Section 4.1.

⁸¹ Chapter 4 will show how this picture of emotions coheres with contemporary philosophy of emotion.

⁸² Bradshaw (2013).

⁸³ Quirke (2014)

the lifespan of a relationship, one that [is] [...] convincing and intimate”.⁸⁴ *La Vie d'Adèle*, then, was widely received as a strikingly accurate and moving representation of love. By representing the two protagonists' relationship in such an intensely emotional way, it seems to *show* what could not be logically argued for given the previously-mentioned deliberative impasse: that if anything constitutes a loving relationship, Adèle and Emma's relationship surely does. Thus, it helps its audience see same-sex relationships as loving relationships, and thereby contributes to generating new shared understandings—here, a more inclusive concept of 'loving relationship'.⁸⁵

A second example indicates how emotionally-laden narrative can publicise new perspectives and conceptual resources during formal public debates. In January 2013, Christiane Taubira, then French Minister for Justice, prefaced parliamentary debates on same-sex marriage with a speech describing marriage's historical evolution. As she reports with a mixture of indignation, resolution, and pride, though marriage in France was initially reserved for Catholics, it then became available to Jews and Muslims. Later, the secularisation of marriage made it available to couples of all faiths, of no faith, and to some hitherto-excluded professions. In parallel, the relationship marriage defines gradually became more egalitarian. Divorce became legal, wives no longer needed their husband's legal authorisation to open a bank account, and, eventually, marital rape was recognised.

By telling this story with passionate rhetoric, Taubira communicated a distinctive perspective on marriage, and this perspective encouraged her audience to reconceptualise marriage as follows. First, according to the perspective she offers, marriage has no fixed meaning. Second, its meaning has always shifted towards greater inclusiveness and equality, and these shifts were widely regarded as moral progress. Finally, with this new conception of marriage in place, extending the right to marry to same-sex couples could appear as the natural next step in marriage's laudable march towards greater inclusiveness. By publicising this conception of marriage, Taubira's emotionally-charged narrative arguably made it easier subsequently to argue for same-sex marriage by appeal to shared premises.

⁸⁴ Clift (2014).

⁸⁵ *La Vie d'Adèle's* accuracy as a representation of lesbian relationships is not undisputed. E.g., Maroh (2013). Accordingly, I present this as an example of epistemically-effective narrative only in a qualified sense: *against a political background* where many citizens denied the standing of same-sex relationships as loving relationships, the film supplied 1) a less inaccurate representation of a same-sex relationship, which 2) vividly showed how unsustainable these prejudices were, and 3) did so in a way that widely impacted public perceptions. This is compatible with retaining doubts regarding its overall accuracy.

In sum, emotionally-charged narrative performs an epistemic function that is distinctively suited to offsetting the overexclusiveness of the Shared Reasons Constraint. The Shared Reasons Constraint's overexclusiveness, recall, results partly from the deep epistemic obstacles diagnosed by social epistemologists. Where different social groups have radically different experiences of the world, and where the conceptual resources needed to make sense of those experiences are often lacking, some considerations that are crucial to publicly contesting injustice will be unacceptable and eventually unintelligible to others. The resulting paucity of shared considerations makes even weaker interpretations of the Shared Reasons Constraint overly exclusive, and limits the extent to which we can introduce new considerations by deriving them from already-shared ones. Emotionally-charged narrative tackles the problem at its source: it facilitates the communication of non-shared experiences and enriches the conceptual resources needed to interpret these. Crucially, moreover, since it is not inferential in form, its ability to publicise new experiences does not—contrary to the 'derivation' strategy—depend on what considerations are already shared. In light of these promising qualities, we should adopt more permissive discursive norms regarding what forms of public discourse are desirable in order to redress the overexclusiveness of public discourse's content. Deliberative ideals that exclude or ignore emotionally-charged narratives deprive themselves of a key epistemic instrument for publicising new considerations, and thereby help to sustain the Shared Reasons Constraint's exclusiveness.

Notice a further advantage of the present proposal. Expanding the forms of deliberation helps to *weaken* the trade-off between the freedom as non-domination of the unreasonable and that of vulnerable participants. On the one hand, it maintains the Shared Reasons Constraint without necessarily excluding the unreasonable, and so preserves its generic freedom-related benefits. On the other, by making the introduction of new shared reasons less dependent on what reasons are already shared, it facilitates the introduction by vulnerable groups of previously-overlooked concerns. Hence, it promotes the non-domination of vulnerable parties, and thereby blunts the force of the overexclusiveness objection. This stands in contrast to the strategy of idealising the Shared Reasons Constraint's constituency. As 3.2 explained, this strategy proposes to reduce the Shared Reasons Constraint's exclusiveness and thus to protect

vulnerable groups from domination, but only by exposing unreasonable citizens to greater risks of domination.

While defenders of the Shared Reasons Constraint often do not give a central role to emotionally-charged and non-argumentative forms of speech in their deliberative ideals, I noted earlier that this is not always true. The rich account of public deliberation Krause offers in *Civil Passions* also combines the Shared Reasons Constraint with an endorsement of emotionally-charged speech in deliberative politics. By way of clarification, then, it is worth noting two residual differences between her argument and my own. First, as we saw in 3.2, Krause readily recommends that we idealise or restrict the perspectives that are relevant to determining which concerns are publicly admissible. I have argued that doing so is problematic: not only does it expose unreasonable citizens to domination, but it may be counterproductive in non-ideal contexts of power. As far as possible, then, we should be cautious about idealising the Shared Reasons Constraint's constituency.

Second, when responding to the overexclusiveness concern, Krause sometimes—though not always⁸⁶—emphasises different elements of her account of deliberation than I do. As 3.3 discussed, she occasionally suggests that what prevents the Shared Reasons Constraint from being too exclusive is the fact that it only governs formal deliberative sites. While I agree that the Shared Reasons Constraint should only regulate formal spheres, I have argued that this is not what helps address the problem of overexclusiveness: distinguishing between different sites in the deliberative system ultimately reproduces the difficulties involved in deriving new considerations from already-shared ones. On my argument, it is rather the inclusion of emotionally-charged narrative which, by publicising the experiences and concerns of divided groups in both formal and informal deliberation, plays a fundamental role in making a relatively non-idealised Shared Reasons Constraint more palatable. Again: by helping to enlarge the pool of shared reasons, it reduces the Shared Reasons Constraint's exclusionary character.

⁸⁶ As mentioned in 4.1, Krause (2008, 162-165) also sometimes emphasises, as I do, the epistemic value of emotionally-charged narrative. See note 78, this chapter.

4.3. Preliminary Objections

Chapter 4 will further elaborate on the function of emotionally-charged narrative in public deliberation, particularly by expanding significantly on the sense in which communicating emotions is epistemically valuable. In the meantime, let us forestall some preliminary worries. To begin, one might object, as Benhabib does, that inviting emotionally-charged narrative into formal and informal deliberation undermines the impartiality of political processes. Emotionally-laden narrative involves conveying one's partial perspective on the world, trying to make one's particular concerns publicly recognised. Hence, doesn't admitting such discourse replace impartial rule with the pursuit of particular interests?⁸⁷

Some philosophers might respond by challenging the value of impartiality.⁸⁸ But the introduction of partial emotionally-charged perspectives is not genuinely opposed to impartiality. Firstly, the discursive norm I am advancing invites partial narratives into public discourse not directly as justificatory grounds for coercively-backed policies, but rather as an epistemic instrument for determining what *should be admissible* as a ground for policy-making. When someone delivers an emotionally-charged narrative, she is not directly making a political demand on the basis of a non-shared worldview. Instead, she is trying to make others see a consideration as a reason for political action. Only if this attempt at making the consideration shared succeeds may that reason be used to make a political demand. While Taubira may deliver an impassioned and partial narrative in legislative debates to get her interlocutors to see marriage in a certain light, she must resort to shared reasons when those debates reach their final stages. So allowing partial modes of expression in this epistemic role is compatible with requiring all coercively-backed policies to be justified impartially, by appeal to shared reasons.

In fact, and secondly, allowing partial narratives is not simply consistent with impartiality: it is also instrumentally necessary for truly impartial public deliberation, which gives fair consideration to the interests of all. Because we occupy particular social positions, no single person reflecting on justice can fairly consider the concerns of all.⁸⁹ Indeed, I unavoidably focus on some considerations to the exclusion of others simply because they are more salient to me, from my particular standpoint. To correct these

⁸⁷ Benhabib (1996, 83).

⁸⁸ Young (1996, ch. 4).

⁸⁹ Krause (2008, 114-115); Young (2000, 108-120).

socially-caused biases, we must engage with occupants of different social positions. In this light, inviting partial narratives helps, rather than hinders, impartiality: by pooling different social perspectives, it helps us register and appreciate the concerns of different social groups.

A more daunting concern is that complementing the Shared Reasons Constraint with a discursive norm assigning a central role to emotionally-charged narrative might seem to overachieve—it is *overinclusive*. Alongside admirable emotional narratives like those mentioned in the last section, this norm might—absent further qualification—open the door for extreme or deeply misguided emotional narratives. For some public speakers, revealing their personal worldview might mean expressing deep-seated rage at racial minorities and immigrants. This is of course not just a hypothetical concern. In conditions of widespread ignorance, demagogues routinely use impassioned appeals to arouse misguided sentiments and thereby steer their listeners to unjust ends. Accordingly, Plato was extremely concerned that intoxicating rhetoric might induce the demos to embrace tyrants and tyrannical policies.⁹⁰ But we hardly need to go all the way back to Plato: this concern is highly salient in contemporary divided democracies, where demagogues such as Marine Le Pen or Donald Trump express narratives of fear and anger which they unfittingly direct at vulnerable minorities.

The first thing to note in response to this concern is that it is not specific to emotionally-laden narrative. Dispassionate argument too can be abused. Indeed, one of the most effective forms of sophistry consists in endowing fallacious inferences with the trappings of rational argument or scientific authority.⁹¹ We are all too familiar with how, as the DSM's classification of homosexuality manifested, purported experts sometimes weigh in misleadingly to public debates by showering audiences with esoteric formulas, technical jargon, or deceptive statistics. So appeals to rational argument and expert knowledge too can be misused. And yet, this is not usually considered a sufficient reason to purge the deliberative ideal of argument or expert testimony.

But given how deeply problematic this 'overinclusiveness' concern is, and how pressing it is in actual divided democracies, we should not leave things at this first observation. Instead, we must investigate more closely what the appropriate normative

⁹⁰ Plato (2007[c. 380 BC], 565e-570c).

⁹¹ Chambers (2009, 337).

response to extreme public narratives is. How should we qualify or complement the norm, advanced in this chapter, that democratic public discourse should invite emotionally-charged personal narratives? What, in other words, should we do about public narratives expressing, say, a racist's rage?

There are actually two overinclusiveness concerns here. The worry might be about the *content* expressed by the narrative, the deeply disrespectful racist views that are publicly communicated. Alternatively, it might concern the *emotion* that is being expressed, independently of what that emotion is about. For example, one might think that publicly expressing rage is morally undesirable, irrespective of what that rage is about. More generally, this second worry targets negative emotions like anger or resentment. The next two chapters will be devoted to exploring these two concerns. Chapter 3 will investigate the problem posed by deeply disrespectful or vilifying speech for democratic discourse that aims to promote justice. As for the worry about negative emotions, I explore it at length in Chapter 4, by defending the place of anger in democratic public discourse.

5. Conclusion

The Shared Reasons Constraint has strong *prima facie* appeal: by facilitating processes of political contestation, it advances the freedom, understood as non-domination, of citizens. As social epistemologists have argued, however, contemporary democracies are characterised by deep experiential divides and hermeneutical injustice, as a result of which some citizens fail to accept and even understand crucial concerns of other citizens. In such settings, even the weakest variant of the Shared Reasons Constraint defines the content of public discourse too restrictively. In doing so, it politically disempowers those whose concerns are misrecognised, and threatens their freedom. So, we have a trade-off relating to non-domination: *pro tanto* reasons grounded in non-domination pull both for and against the Shared Reasons Constraint.

The three most influential strategies advanced by defenders of the Shared Reasons Constraint for addressing this problem encounter significant limitations in non-ideal contexts. Deriving novel considerations from already-shared ones is highly difficult if we start from settings where some groups lack the experiences and conceptual resources needed to appreciate others' concerns. To bolster this strategy, one might

restrict the Shared Reasons Constraint's constituency to idealised participants. Doing so expands the pool of initial shared reasons from which the derivation of new reasons then proceeds. However, this proposal is ill-advised in contexts where unreasonable citizens are empowered. And even outside of such contexts, it exposes unreasonable citizens to domination. This strategy therefore does not attempt to *weaken* the above trade-off, but simply settles it in favour of reasonable citizens. Hence, it should only be considered as a last resort. Finally, confining the Shared Reasons Constraint to formal sites in the deliberative system, though appealing on other grounds, only postpones the derivation strategy's problems.

To complement these strategies for mitigating the Shared Reasons Constraint's overexclusiveness, we should adopt more expansive norms concerning the *form* of deliberation. In addition to arguments from shared premises, formal and informal public deliberation should welcome emotionally-charged narratives. In virtue of their non-inferential character and of their emotional tenor, such narratives constitute an effective epistemic tool for publicising new considerations. Therefore, by giving such discursive forms a greater place in democratic public deliberation, we can maintain the Shared Reasons Constraint while attenuating its negative impact on vulnerable groups' political freedom.

Nevertheless, absent qualification, this proposal may seem overinclusive. Firstly, it might open the door for narratives whose content is deeply disrespectful. Secondly, it might invite narratives charged with highly negative emotions, such as anger and rage. Even if one thinks that some emotions have an important part to play in political discourse, one may have reservations about profoundly negative emotions. The next two chapters address these two concerns in turn.

CHAPTER 3

COUNTERING PUBLIC HATE SPEECH

1. Introduction

1.1. The Overinclusiveness Problem

I have been concerned with the following question: what norms should regulate democratic public deliberation in non-ideal conditions marked by pervasive inequalities and profound ethical divisions, for inclusive deliberation to constitute an effective tool for advancing justice?

Chapter 2 introduced the Shared Reasons Constraint, according to which participants must appeal to considerations that are shared by others. This norm is *prima facie* appealing, as it facilitates processes of contestation whereby participants hold powerful parties accountable. However, one of the principal objections to this constraint is that it tends to be overly exclusive in non-ideal contexts: it excludes some important considerations, without which some participants cannot effectively pursue justice.

To remedy this deficiency while maintaining as far as possible the Shared Reasons Constraint's advantages, I recommended that we expand our conception of what forms of discourse are publicly admissible. In particular, normative models of democratic public deliberation should welcome emotionally-charged narratives as epistemically useful. By communicating particular experiences and broader worldviews, such narratives contribute to making previously excluded considerations shared. If this perspective-sharing process is successful, the Shared Reasons Constraint *then* allows the newly-shared considerations to be used when justifying one's political demands.

While making space for personal narratives in this way helps tackle the problem of overexclusiveness, we saw at the end of the last chapter that it generates the opposite problem—a problem of overinclusiveness—which constitutes the present chapter's focus. In strongly divided democracies, some citizens have highly prejudiced beliefs about others. For such citizens, telling others how they see the world may involve publicly expressing deeply disrespectful views—i.e., views that deny others' standing as free and equal persons. Racists or homophobes, for example, might publicly deliver narratives that communicate their hateful viewpoints. So, disrespectful views could

enter public discourse at the deliberative stage where citizens express narratives aiming to make their distinctive perspectives shared.¹ Unless they are further qualified or supplemented, then, the norms advanced in the last chapter may open the door for public communications whose content is abhorrent.

That deeply disrespectful views may make their way into public discourse is of course not just a theoretical problem. Whether it is Danish cartoonists publishing representations of the prophet Muhammad as a bomb-throwing terrorist,² British journalists assimilating migrants to “cockroaches” and to a “plague of feral humans”,³ or then-presidential candidate Donald Trump referring to Mexican immigrants as “rapists” and “criminals”,⁴ one thing leaves no doubt: disrespectful and degrading speech is rampant in actual democratic political discourse.

What should we do, then, about public speech that expresses deeply unjust and disrespectful viewpoints? There is no real controversy that such utterances are morally undesirable. Deliberative theorists of democracy widely agree that, ideally, participants in deliberation should respect one another. Gutmann and Thompson, for example, maintain that “mutual respect” should govern deliberative interactions.⁵ Likewise, in Young’s communicative ideal, interlocutors should frame their exchanges “in terms of basic respect”.⁶

Thus, by contrast with the discursive norms mentioned in Chapter 2,⁷ the salient question here is not whether deliberators have moral reasons to refrain from engaging in deeply disrespectful speech. That much is uncontroversial. Rather, the salient question is: given that so many actual speakers violate this moral norm, what kind of normative response is warranted by these moral reasons? Do the moral reasons that oppose disrespectful public speech recommend implementing formal legal norms that coercively expel such speech from democratic public deliberation? Or should we

¹ This is true even if the narrative ultimately fails to make the abhorrent considerations shared. While such a failure would mean that the Shared Reasons Constraint will not allow the consideration to be used to justify political demands during the final stages of formal deliberation, the abhorrent narrative would still be expressed at the prior deliberative stage where deliberators aim to publicise their distinctive experiences.

² Asser (2010).

³ Osborne (2015).

⁴ Reilly (2016).

⁵ Gutmann and Thompson (2004, 79-81).

⁶ Young (2000, 48). See also Krause (2008, 166) and Mansbridge et al (2010).

⁷ As we saw in Chapter 2 (esp. Section 2.1), debates about the Shared Reasons Constraint are largely about whether there *is* a moral duty to appeal to shared considerations, where this moral duty might be unenforced or only enforced informally (e.g., through verbal disapproval).

instead respond to disrespectful speech internally to democratic discursive processes, by countering it with more speech?

My discussion will focus on an especially abhorrent type of disrespectful speech: hate speech. There are two reasons for this focus. First, hate speech picks out an intensely disrespectful kind of speech, as opposed to a mildly disrespectful one. Consequently, the public presence of hate speech is particularly threatening for the view that inclusive public deliberation can promote justice in non-ideal conditions. Second, philosophers of language, feminist and critical race theorists, and legal philosophers have closely examined hate speech and have developed rich conceptual resources for understanding the injustices it may enact. Democratic theorists concerned with making the deliberative ideal relevant to divided settings have much to gain from incorporating these important theoretical insights.

1.2. Defining Hate Speech

The definition of ‘hate speech’ is highly contested. By ‘hate speech’, I will refer, roughly, to communications that emphatically deny the basic status of other members of society as free and equal citizens. More specifically, I follow influential discussions in taking hate speech to deny the basic standing of individuals who belong to vulnerable social groups, in virtue of their belonging to these groups.⁸ Two things are worth stressing here. First, hate speech includes not just verbal speech, but also all platforms (including printed images) that might be used to communicate propositional contents. Second, what makes a communication hate speech is not the speaker’s emotion, but rather the content of the communication. Chapter 4, however, *will* examine the place of a controversial emotion, anger, in public deliberation.

This abstract characterisation is intended as an approximate guide, rather than as a precise analysis. So, to make matters clearer, consider a few illustrations of the kinds of speech I am concerned with. Hate speech might include newspaper articles falsely attributing essential dangerousness to a minority group (‘Muslims are terrorists’) or leaflets portraying some groups as subhuman (for instance, depictions of blacks as ape-like). Sometimes, hate speech explicitly expresses the exclusion of a minority group

⁸ This definition is inspired by Waldron (2012, 27-28, 34-41) and Brettschneider (2012, 1). In legal scholarship, it echoes the UN’s (1965) condemnation of speech “based on ideas or theories of superiority of one race or group of persons of one colour”.

from the political community, as in Pegida's infamous posters asserting 'Rapefugees Not Welcome'.

My discussion will focus on hate speech occurring in *public discourse* within *stable democracies*. The second restriction brackets contexts of political instability where hate speech is likely to induce a breakdown of minimal peace. While this would constitute an extremely powerful reason for legally banning it,⁹ I will rather be concerned with cases, like France or the US, where hate speech is diffuse without threatening to produce such an imminent political breakdown.

As for the first restriction, I understand 'public' quite broadly.¹⁰ It encompasses communications occurring in formal political debates, but also those published in newspapers, posted on the internet, or plastered in common spaces. Very often, Waldron notes, such communications become "a permanent or semipermanent part of the visible environment".¹¹ Moreover, this is commonly speech that is not simply directed at the targets of the hate speech's content, but addresses a broad audience, which includes third parties. Hence, although the distinction between public and non-public hate speech remains vague, this restriction tends to exclude live, face-to-face encounters where, for example, the speaker directs a slur at a particular person with the deliberate intention to wound, or cases where someone threatens a particular individual by, say, spray-painting a swastika on their door.¹²

I focus on *public* hate speech, thus defined, for two reasons. The first is dialectical. I will chiefly be engaging with Jeremy Waldron, and he takes his argument for banning hate speech to be most powerful with respect to such public hate speech. The second is political. The political contrast that is often drawn is between the US, which tends to oppose hate speech bans, and the rest of the world, which tends to embrace them. But even in the US, live, face-to-face "fighting words", as well as targeted threats and harassment, are generally legally prohibited.¹³ So the main debate concerns not *whether*, but *which* kinds of hate speech should be prohibited. And hate speech bans in public discourse are politically far more divisive.

⁹ Tirrell (2012).

¹⁰ See also Chapter 1, note 40.

¹¹ Waldron (2012, 37).

¹² Heinze (2013, 590-591), Langton (2012, 76-77), Brettschneider (2012, 1), and Waldron (2012, 37) also bracket such face-to-face assaultive forms of hate speech. Brison (1998, 313-315) treats them separately.

¹³ *Chaplinsky v. New Hampshire* (1942); *Virginia v. Black* (2003).

This characterisation might still seem too broad. Many actual hate speech regulations require not just that the expressed content of the regulated speech profoundly contradict the basic standing of vulnerable groups, but also that it be uttered in circumstances such that it is likely to produce certain effects, like violence or antipathy.¹⁴ Now, there is enormous variation in characterisations of hate speech. In part, this is due to the fact that how we define hate speech depends on what kinds of speech-related harms we are concerned with. Accordingly, I adopt this relatively broad account because 1) it is adapted to the particular speech-related harm I will be focusing on (see Section 2), and 2) it is informative given my more general project: since I am fundamentally interested in democratic deliberation, I am especially concerned with speech that is abhorrent because it expresses views that contradict the moral underpinnings of democracy (i.e., participants' status as equal and free citizens).

1.3. Outline

What should be done about hate speech occurring in public discourse within stable democracies? Should we supplement the norms defended in Chapter 2 with a further constraint, by coercively banning hateful considerations—that is, should legal norms threaten to impose penalties on those who publicly engage in hate speech? Or should we tolerate such speech, and try to address the injustice it enacts internally to democratic discursive processes?

The relevant literature is vast, and spans across philosophical and non-philosophical domains. Consequently, in addressing these questions, I will predominantly engage with Waldron's treatment of hate speech. Waldron offers a philosophical analysis of what makes public hate speech harmful, which he believes should incline us to legally suppress such speech. Taking Waldron's framework as our starting point is fruitful for three reasons. First, Waldron's argument is clear and powerful: not only is it rooted in an appealing picture of justice and of what it means to respect human dignity, but—in line with my aim of theorising democratic public deliberation for non-ideal conditions—it is acutely sensitive to how deeply harmful hate speech can be. Second, and as a result, Waldron's framework has been highly influential in contemporary

¹⁴ The UK's Public Order Act of 1986, for instance, prohibits insulting or abusive speech when it is *likely to stir up hatred*. See also Mill's (2006[1859], 64-65) justification for restricting speech.

theorising about hate speech. Finally, it draws on a wide range of sources, and therefore helpfully integrates the insights which law, political philosophy, feminist thought, and critical race theory have contributed to the issue.

The principal strength of Waldron's analysis lies in its articulation of how public hate speech harms the dignity of its targets.¹⁵ There is no denying, therefore, that if unchallenged, public hate speech is deeply harmful and unjust. But what is less persuasive, I will suggest, is the claim that coercive legal bans constitute the best way of responding to this harm. Waldron, like many other proponents of bans, relies on an overly optimistic picture of his preferred solution (bans), and an overly unsophisticated picture of the main alternative (countering hate speech with more speech). By offering a more refined understanding of 'counterspeech', and a more nuanced vision of bans, I will argue that critiques of counterspeech either fail or apply even more forcefully to bans. In this light, we should be significantly less inclined to implement legal norms banning public hate speech. Instead, we should embrace robust norms of democratic counterspeech.

Note, however, two limitations of my argument in this chapter. First, my defence of counterspeech remains pessimistic. In other words, it will proceed less by suggesting that democratic counterspeech fully overcomes the objections levelled at it, and more by showing that these objections are at least as problematic for bans.

Second, although I focus primarily on Waldron's influential account of what makes public hate speech harmful, there are other ways in which such speech can harm. For this reason, my argument in this chapter is limited in scope. That being said, it does not follow that my inquiry is irrelevant to these other harms: as a further line of investigation, we should explore to what extent the strategy I deploy for defending counterspeech can successfully be extended to other speech-based harms. Because many of the other harms associated with hate speech involve claims about the causal effects of hate speech, I expect that this further line of investigation will be in substantial part empirical.

My argument will be structured as follows. I begin by outlining Waldron's framework for understanding the unjust harm enacted by hate speech, and why he takes it to support legal bans (Section 2). In turn, I critique this defence of bans by articulating a more sophisticated understanding of democratic counterspeech, and why

¹⁵ For criticism, however, see Simpson (2013).

it seems a better way of blocking the harm in hate speech (Section 3). Now, the policy of relying on such counterspeech does face genuine difficulties. Nevertheless, I then demonstrate that the two weightiest objections to counterspeech can be turned, with greater force, against bans (Section 4-5). Section 6 concludes.

2. The Dignitarian Case for Bans

For some, the harm involved in hate speech warrants forcibly excluding or banning it from the public deliberation of democratic societies.¹⁶ According to their opponents, countervailing considerations, such as freedom, democracy, or truth, demand that we tolerate it.¹⁷ Waldron articulates a framework for understanding the harm in hate speech which, he claims, supports the former option.

Waldron starts from an appealing picture of justice (a picture consonant with the one that informs my broader inquiry¹⁸) whereby justice requires treating human beings with dignity. In the first place, this means that citizens' status as equal and free members of society in good standing—their dignity—should be recognised and upheld by other citizens.¹⁹ However, drawing on Rawls's idea of a well-ordered society, Waldron argues that justice requires something more. Citizens must also *know* that their peers uphold their good standing. Absent the assurance that their dignity is safe, citizens cannot fully enjoy their good standing. Indeed, this assurance is necessary for citizens to pursue their aims and participate in civil and political life without fear or shame.²⁰

Because the public assurance of dignity is an essential component of justice, Waldron suggests, we should pay greater attention to “political aesthetics”: we should care, in other words, about what our society literally looks and sounds like.²¹ Imagine, Waldron continues, what a hate speech-allowing society would look like to targets of hate speech. Think of the “marching feet and chants of neo-Nazis in Skokie, a Grand

¹⁶ E.g., Waldron (2012), Delgado (1993), Matsuda (1993), Brison (1998), and Quong (2011, 305-312).

¹⁷ On freedom or autonomy, see Baker (1996), Scanlon (1972), and Dworkin (1981a). On democracy, see Heinze (2016), Dworkin (2009), and Post (1990). On truth, see Mill (2006[1859]).

¹⁸ As discussed in Chapter 1, Section 2, the idea of justice that informs my project involves upholding social relations of equality. This is consonant with Waldron's dignitarian picture of justice, since Waldron defines 'dignity' as the social status of being an equal member of society.

¹⁹ Waldron (2012, 59-60).

²⁰ Ibid., 65-69, 81-89.

²¹ Ibid., 71-77.

Wizard's speech at a Ku Klux Klan rally".²² Think of a Muslim citizen confronted with newspapers depicting Muslims as terrorists. This is clearly not what a well-ordered and just society looks and sounds like. When a society looks like this, it is not public knowledge that all citizens uphold the dignity of others.

Hence, we can appreciate a powerful harm generated by public hate speech. In publicly denying the status of its targets as social equals, hate speech affects the way the society looks in such a way as to undermine the public assurance that their status is secure. Moreover, public hate speech reaches out to other hateful persons, to assure them that they are not alone. In doing so, it replaces the assurance of dignity with the assurance of hatred.²³

How exactly does hate speech relate to this harm? Dignitarian accounts of the harm in hate speech commonly argue that hate speech *causes* dignity-related harms. On this understanding, hate speech causally influences how other members of society view and act towards the target group, so that they stop upholding their status as equals.²⁴ While Waldron's account is consistent with the causal claim, he largely focuses on how hate speech *constitutes* an assault on dignity.²⁵ In publicly saying 'You are not one of us, and we want you out', hate speech undermines the public assurance that the target group are respected by their peers. Communicating this is part and parcel of the act of refuting the public assurance of dignity, and of assuring other citizens with hateful views that they are not alone.²⁶

For Waldron, the fact that hate speech constitutes an assault on its targets' effective enjoyment of dignity reveals how fragile this good is. "The flare-up of a few particular incidents can have a disproportionate effect" on the public assurance of dignity.²⁷ Hearing racist vituperations of far-right groups on the radio may well produce fear and a sense of alienation in its targets, even if the views uttered are marginal. This, Waldron concludes, is why coercive hate speech bans are required. By forcefully excluding hate speech from the public sphere, bans prevent such speech from creating an environment of fear and alienation.²⁸

²² Ibid., 71.

²³ Ibid., 94.

²⁴ E.g., Delgado and Stefancic (2004).

²⁵ Waldron (2012, 166-167).

²⁶ For other constitutive accounts, see, e.g., Langton (2012) and Maitra (2012).

²⁷ Waldron (2012, 94).

²⁸ Ibid., 93-95.

This constitutes the basic structure of Waldron's argument for banning public hate speech. As I critically evaluate it, I will progressively introduce further complexities and nuances. In the meantime, notice three features of this argument that are particularly appealing. Firstly, it rests on intuitively appealing moral foundations—the idea that justice requires not just respecting the dignity of human beings by upholding their status as equal members of society, but also publicly assuring them that their equal status is secure.

Secondly, the framework is methodologically appealing. Waldron embraces non-ideal theory by demanding that we imagine what hate speech looks like in the “real world”, particularly to those on the receiving end. This commitment to taking the experiences of targets seriously is a welcome corrective to scholarship which, he suggests, frequently adopts an abstract and idealised conception of hate speech.²⁹

Finally, because it focuses on the harms constituted (as opposed to caused) by hate speech, Waldron's framework is dialectically powerful. It sidesteps the fiercely contested empirical question of whether the causal relations between hate speech and its purported harms really obtain.³⁰

Consequently, Waldron's philosophical framework constitutes a promising starting point for thinking about what we should do about public hate speech. But it remains problematic. While the account of hate speech's harm is convincing, his inference to the endorsement of hate speech bans is too direct, and ultimately unpersuasive.

3. A Fairer Comparison

Public hate speech assaults the dignity of its targets. How can we uphold justice and (as a component thereof) the assurance of dignity in the face of this assault? While Waldron advocates banning public hate speech, an alternative would be to allow such speech, and to counter it with more democratic public speech. Implicit in Waldron's above endorsement of hate speech bans, then, is the belief that such democratic counterspeech is insufficient. Unless democratic public discourse is already purged of hate speech, it constitutes an inadequate instrument for promoting justice and the

²⁹ Ibid., 33.

³⁰ Ibid., 166.

assurance of dignity. So, if we care about using inclusive public deliberation to promote justice in real-world conditions, hate speech bans are needed.

The central problem with this inference to hate speech bans is that it relies on an unfair comparison between the society that allows hate speech and the society that bans it: as we will see, Waldron is overly idealistic about the latter and overly uncharitable about the former. The present section aims to rectify this unfairness. Take the hate-banning society first.

The Hate-Banning Society: The society in which bans suppress public occurrences of hate speech by imposing coercive sanctions on such speech.

The most commonly invoked worry with this society is that having one's speech suppressed significantly undercuts one's freedom or autonomy. Although this point can be made with different conceptions of autonomy or freedom, I will emphasise, in line with previous chapters, how it applies with freedom as non-domination. Domination, recall, is subjection to alien or unchecked power. Accordingly, non-domination is secured insofar as individuals are capable of contesting the power to which they are subjected, so that it is forced to track their interests and concerns.³¹ In Chapter 2, I argued that the ability of participants to contest and thereby check political power is enhanced when they can publicly appeal to the considerations they deem most significant.³² This remains true, the present worry continues, even when those considerations are hateful ones. Hence, banning public hateful considerations diminishes the freedom as non-domination of citizens with hateful political views.³³

Importantly, Waldron himself recognises these freedom- or autonomy-based costs. Bans, he claims,

have a direct bearing on freedom to publish, sometimes on freedom of the press, very likely on freedom of the Internet [...] And this matters to

³¹ Lovett and Pettit (2009, 14).

³² Chapter 2, Section 2.2.

³³ For other freedom- or autonomy-related objections, see Brettschneider (2012, 76-80), Scanlon (1972), Dworkin (1981a), and Baker (1996, 992). Note that this objection cannot justify protecting all kinds of hate speech. Shouting a slur at someone in the street, for instance, is not an attempt at contesting political power. But, as stated in the introduction, I am bracketing such assaultive, face-to-face forms of hate speech. By contrast, the examples of public hate speech mentioned in the introduction (e.g., 'Muslims are terrorists') sometimes *are* raised to support political demands.

individuals. Often the messages that racists or Islamophobes are stopped from expressing are the very messages [...] that matter most to them. For them, other aspects of political expression pale into insignificance [by comparison]³⁴

But he responds that these considerations are outweighed by considerations having to do with the dignity of targets of hate speech. All things considered, then, bans remain morally justified.

We could go even further on Waldron's behalf. As Susan Brison has argued, lacking assurances that one is a citizen in good standing risks depriving one of the psychological conditions needed to participate politically without fear or shame.³⁵ By undermining the public assurance of dignity, hate speech may therefore undercut its targets' ability to contest exercises of political power. Therefore, even *within* the value of freedom (understood as non-domination), one might think that the balance of reasons favours bans. I want to grant these responses to the freedom-based argument, not just for the sake of argument, but also because—unlike some critics³⁶—I find such claims about the harmfulness of hate speech persuasive.

Even so, we should still be worried about bans. Although the freedom-based reasons against bans may be outweighed, they remind us that upholding the dignity and freedom of targets of hate speech comes at a significant price. Part of what is non-ideal in actual circumstances is that many citizens uphold hateful viewpoints. Witness the rise of European far-right parties like the French *Front National*, whose rhetoric is often saturated with xenophobia. In some cases, moreover, supporters of such hateful political ideologies are socioeconomically disaffected citizens who feel, sometimes justifiably, that their voices have been ignored by mainstream parties.³⁷ Thus, coercively suppressing hate speech seems to come at a substantial moral cost: it further diminishes the freedom of many relatively disadvantaged citizens, and thereby fuels the exclusionary processes that, often, contributed to producing their hateful disaffection.³⁸

³⁴ Waldron (2012, 148-149).

³⁵ Brison (1998).

³⁶ E.g., Dworkin (2009, vi).

³⁷ E.g., Willsher (2015).

³⁸ Why does it matter that those affected by bans are often relatively badly off? On a consequentialist interpretation, ignoring those who have already been ignored is problematic because it risks alienating them further, and breeding more hatred. But the explanation needn't be consequentialist. Instead, one might think

What this means is that even if Waldron is right that the claims of targets of hate speech outweigh those of citizens who possess and wish to express abhorrent views ('hateful citizens', for short), the trade-off between the two is significantly worse than he acknowledges. This realisation should strongly motivate us to seek a solution that weakens the trade-off—that is, that blocks the harmful features of hate speech *while* limiting the costs to hateful citizens. Bans should be a last resort.

Now, in observing that many of those whose freedom would be restricted by bans are relatively disadvantaged and numerous, my intention is to bring out how important it is to find an alternative response to hate speech. But my more fundamental point does not hinge on this sociological observation. Even if hateful citizens tended to be well off (as they of course often are), there would still be *a* freedom-based cost to restricting their political speech. Accordingly, it would remain true that *if* we can find an alternative solution that blocks the harm in hate speech without imposing this cost on hateful citizens, we should prefer that solution to legal bans.

To this end, let us now consider the society that allows hate speech, legally speaking. Waldron's most sustained depiction of this society is uncharitable. Opponents of hate speech bans "love the richness and untidiness of the marketplace of ideas: let a thousand flowers bloom, they say, even the poisonous ones".³⁹ On this crude understanding, the hate-allowing society responds to the interests of hateful citizens by allowing them to publicly disclose their abhorrent worldviews. As for the interests of targets of hate speech, the idea is that since *all* citizens can voice their opinions, hateful views can be countered by more public speech, which challenges hateful utterances.

Critics have forcefully argued that this counterspeech-based solution to the problem of hate speech is profoundly inadequate, for two reasons. First, it seems unfair, insofar as it typically places the burden of challenging hate speech on its targets.⁴⁰ Second, it seems ineffective at protecting targets of degrading speech. To begin, empirical evidence indicates that targets of hate speech rarely actually speak back in public. This suggests that the harmfulness of hate speech makes speaking back very difficult.⁴¹ Moreover, even if targets of hate speech did speak back, it is unclear how effective this

this matters because one shares the basic prioritarian intuition that losses of welfare or freedom are more morally problematic when they afflict those with overall lower levels of well-being.

³⁹ Waldron (2012, 67). Nielsen (2012, 154), Brown (2015, 246), Matsuda (1993, 36-37), Delgado (1993, 95), and MacKinnon (1993, 72-73) also criticise this unrefined understanding of counterspeech.

⁴⁰ Maitra and McGowan (2012, 9).

⁴¹ Nielsen (2012).

would be at preserving the public assurance of dignity. The difficulty, as philosophers of language often put it, is that targets of hate speech frequently lack the *authority* needed to block the harm in hate speech.⁴² If what is harmful in public hate speech is that it asserts that a given group are unwanted by the society at large, then being able to refute this assertion requires being able to speak on behalf of the society at large. Since targets of hate speech are typically members of minority groups, they are often not in a position to do this.

However, this picture of the hate-allowing society and of the democratic counterspeech it involves is largely⁴³ a straw man. Setting up a choice between either having the state ban hate speech or having the state step back and let private citizens respond is misleading. Instead, as Corey Brettschneider has argued, the state can be active as a speaker endowed with expressive powers. So a more refined alternative to bans is the following:

The Hate-Allowing Society: The society where hate speech is not coercively suppressed, but where the state actively speaks out against public occurrences of hate speech.⁴⁴

States possess numerous tools for speaking back. To begin, the state should speak back by having public officials affirm the ideal of human dignity, not just through public pronouncements in formal deliberative forums, but also by naming public spaces, dedicating monuments, and enacting public holidays. Moreover, state officials should speak back by denouncing particular instances of public hate speech.

How does this state-based counterspeech address the above objections? First, while this solution is consistent with the victims of hate speech also speaking back, it is not primarily victims who are required to respond. Rather, it is public officials or, as I explain below, third parties. Thus, this solution seems less unfair.

Second, state officials, unlike many private citizens, have an authoritative status, in virtue of which they can successfully challenge hate speech's attempted assault on the public assurance of dignity. In decrying detestable speech, the state is *reassuring* its

⁴² Maitra and McGowan (2012, 9-10).

⁴³ Some advocates of counterspeech do endorse the unsophisticated version. E.g., Rauch (1993, 159), Richards (1999, 135-136), Schulzke (2016, 237-239).

⁴⁴ Brettschneider (2012).

targets. It is saying: ‘We are sorry that you had to hear that. It’s true, some individuals unfortunately still believe this. But *on behalf of the people generally*, rest assured that we stand by you, that they, not you, are the outliers, and that we will ensure that what they say remains an empty threat.’

To make this more vivid, consider how Christiane Taubira—then the French Minister for Justice—intervened in response to David Rachline—a *Front National* Senator—in the French Senate in 2015. After Rachline criticised Taubira’s advocacy of humane treatment for criminals, whom he referred to as “thugs” and “executioners”, Taubira replied to his various claims, and concluded as follows: “You belong to a political family [...] that systematically seeks scapegoats [...] and that looks for invaders everywhere [...] [O]ne shudders at the thought of what would happen if it took power—but we will fight to make that absolutely impossible”.⁴⁵ The overwhelming majority of the Senate rose to their feet to applaud this response. In her response, Taubira does several things: she takes the time to answer; she denounces far-right ideologies and their scapegoating practices; she promises to keep those exclusionary ideas away from power; and she does this in broad daylight, to wide acclaim. If we want hate speakers isolated, so that they cannot undermine the assurance of dignity, public counterspeech like Taubira’s seems an exemplary realisation of those aims. Her response not only contributes to making Rachline’s ideology isolated (as a ban might achieve), but also publicly shows it to be isolated.

Of course, public officials can only do so much. But the state can also speak back in more indirect ways. First, as Brettschneider notes, it can fund private associations, like civil rights groups, whose functions include publicly denouncing occurrences of hate speech. This provides support to targets of hate speech by empowering third parties to block its attempted harms.⁴⁶ Second, the state can speak against hateful views through public education, which teaches children about historical events like the Holocaust, and which exposes them to democratic values.⁴⁷

Now, the latter recommendation involves state coercion as well as state speech. Parents are required to send their children to school. Consequently, they will almost unavoidably have to come into contact with worldviews they disavow. Nevertheless, it

⁴⁵ Sénat Vidéo (2015).

⁴⁶ Brettschneider (2012, 110-111).

⁴⁷ Ibid., 96-104.

does not prevent hateful parents from publicly expressing their viewpoints.⁴⁸ And it is coercively suppressing such expressions that is especially problematic for individual freedom. Recall: doing so is problematic, in part, because it reduces one's ability to contest exercises of political power. By contrast, having to send one's children to school, or hear other viewpoints, does not.

Admittedly, however, countering hate speech through public education does entail that teachers, in their capacity as teachers, are required to refrain from expressing hateful viewpoints. As Brettschneider notes, "there is no general right to be employed by a public school". Teachers (like other government employees) are hired *on the condition* that they perform certain functions—including repudiating hateful views in the classroom. Therefore, the state would be "justif[ied] [in] firing teachers who spread such views".⁴⁹

Let us take stock. On this refined picture of democratic counterspeech and of the hate-allowing society, the state intervenes to block the dignitarian harms of hate speech. But it does so without coercively suppressing hateful citizens' public utterances. Hence, it weakens the aforementioned trade-off between the freedom of hateful citizens and the freedom of their targets. In light of this fairer comparison between the hate-allowing society and the hate-banning society, the former seems *prima facie* more desirable. It is, in other words, a better way of upholding justice and the public assurance of dignity in the face of public hate speech. In what follows, I consider two objections which suggest that my argument thus far idealises state-sponsored counterspeech. In both cases, I will show that counterspeech and bans are companions-in-guilt: these objections can be turned—with greater force—against bans.

4. Objection 1. Does Counterspeech Block All Hate Speech?

4.1. The Limited Effectiveness of Counterspeech

One might think that it is unrealistic to hold that state-based counterspeech can truly block the dignitarian harm in hate speech. In his very brief consideration of state-based counterspeech, Waldron objects that this solution would provide the wrong kind of public assurance of dignity. In particular, raising "billboards proclaiming the principles

⁴⁸ Ibid.

⁴⁹ Ibid., 102.

of justice” would have a “creepy totalitarian flavour”.⁵⁰ To avoid this, the assurance of dignity should be provided in a way that is more implicit.

This response seems misguided in several respects. To begin, it is unclear in what sense laws banning hate speech constitute implicit assurances. After all, such laws must be publicly known and typically lead to widely-publicised lawsuits. But even granting this, we can query the claim that explicit assurances are undesirable. First, Waldron makes a straw man of speaking back. As the Taubira example illustrated, counterspeech often involves public responses to particular instances of hate speech that are not flat-footed in the way Waldron’s envisaged billboards are. Moreover, even if state counterspeech does involve explicit pronouncements of values, is that really a bad thing? Does Waldron really think that billboards—like those displayed in football matches—stating ‘SAY NO TO RACISM’ are “creepy”? And is explicitly erecting monuments honouring the Civil Rights Movement “totalitarian”?

Waldron suggests that such explicit assurances are bad because they are “evidence of a problem”.⁵¹ But bigotry *is* a problem. Given that this is the case, it seems at least an open question (as we have seen in Section 3) whether it is preferable to recognise the problem and publicly denounce it, or maintain a blissful ignorance of its existence by suppressing the public speech of hateful citizens.

A better version of the objection under consideration comes from philosophers of language who argue that degrading speech often takes a form that is very difficult to identify and speak out against. Rae Langton, in particular, has emphasised how degrading representations are often *presupposed* rather than straightforwardly asserted in hate speech.⁵² Instead of saying ‘Blacks are lazy’, someone might say ‘Even blacks would do that job’, thereby implying that blacks are lazy.

Why does this matter? Conversations tend to follow rules of accommodation. Very roughly, this means that the content of presuppositions tends to be added by default to the common background assumptions that the relevant conversation relies on, unless those presuppositions are explicitly challenged. Suppose you say ‘I need to walk my dog’. The idea is that it automatically becomes an accepted assumption in our conversation *that you have a dog*, unless I say ‘Wait a minute, you don’t have dog!’⁵³ The

⁵⁰ Waldron (2012, 87). See, similarly, Alexander (2017, 372-373).

⁵¹ Waldron (2012, 87).

⁵² Langton (2012, 82-84).

⁵³ Ibid., 82-89.

accommodation of presuppositions means that some contents can be smuggled in more easily to the conversational common ground than if they were straightforwardly asserted.

On its own, this is not immediately problematic for state-based counterspeech. If the problem is that degrading representations presupposed in hate speech automatically become common ground unless they are blocked, then it seems this rather *calls* for counterspeech. While one way of stopping the degrading representation of black citizens from becoming common ground is to prohibit utterances of ‘Even blacks would do that job’, the other is to block it through counterspeech: ‘What do you mean, *even blacks*?! We don’t condone those views around here!’

It is in light of a further observation, concerning coded speech, that the point about presupposition becomes particularly challenging. Jason Stanley has argued that the conventional meanings of words can shift, so that hateful representations become encoded in ordinary words. The idea is that repeated associations in the media of ordinary terms, like the association of ‘inner cities’ with black citizens, eventually make this term refer to black citizens. Similarly, repeated associations of ‘welfare’ or ‘food stamps’ with laziness make ‘welfare’ or ‘food stamps’ encode laziness.⁵⁴

When combined with the existence of coded speech, Langton’s observation regarding presuppositions becomes highly problematic. For presuppositions not to be added to the conversational common ground, they must be identified and challenged. But doing so is far more difficult if they are concealed through coded speech. More precisely, when degrading contents are encoded in public speech, the message is typically quite clear to its targets and to other hateful citizens, but less visible to third parties. This, in turn, is particularly worrying for *state-based counterspeech* since it emphasises the need for third parties, including public officials, to play a central role in responding to hate speech.

To illustrate, consider Stanley’s example of an interview conducted in 2012 with Newt Gingrich, a leading US politician. Gingrich had previously claimed that “black Americans should demand jobs, not food stamps”. In response, the interviewer observed that this was offensive to blacks, as Gingrich was clearly invoking negative

⁵⁴ Stanley (2015, 123, 138, 156-157).

racial stereotypes—namely, that blacks are lazy. Gingrich answered “No, I don’t see that”, and was loudly applauded, while the interviewer was booed.⁵⁵

This case, Stanley suggests, exemplifies the use of coded presuppositions expressing degrading contents in public speech. Gingrich’s claim presupposes *that black Americans ask for food stamps rather than jobs*. And if ‘food stamps’ encodes laziness, the presupposition means *that black Americans are lazy*. This is problematic, firstly, because it makes identifying the degrading stereotype being expressed more difficult. But the example also reveals a second problem. Even when one has identified the presupposition, speaking back against it risks backfiring. Because the degrading content is encoded, the person who wishes to block it (here, the interviewer) must be the first to explicitly invoke it. This makes it possible for the person who challenges the hateful content to be blamed for its expression, so that the intervention achieves the opposite of its intended effect. Instead of assuring black citizens of their standing, the would-be challenger is made to seem as though he holds degrading views.

What is worse, the above example is hardly the most coded imaginable. Substituting ‘inner cities’ for ‘black Americans’, Gingrich could have presupposed the same demeaning content by saying ‘Inner cities’ reliance on food stamps is draining public resources’. Stanley goes even further, and suggests that the very words ‘welfare’ or ‘food stamps’ can sometimes immediately encode *that blacks are lazy*, even in utterances that do not appear to be about black citizens (e.g., ‘Food stamps are expensive’). In these cases, what the terms presuppose is their own conventional meaning, which has been altered to encode the degrading stereotype.⁵⁶ Hence, the difficulties illustrated in this example might easily be amplified. In sum, coded speech, which Stanley argues is widespread, seems to generate profound difficulties for state-sponsored counterspeech.

4.2. Two Replies: A Defence and an Attack

A first reply is that we should nonetheless not conclude too much from these difficulties. Stanley sometimes draws deeply sceptical conclusions regarding counterspeech from the foregoing analysis. He suggests, for instance, that the contents

⁵⁵ Ibid., 156-167.

⁵⁶ Ibid., 138.

encoded in speech are “not negotiable”.⁵⁷ So, the existence of coded presuppositions renders effectively speaking back through democratic deliberation “impossible”.⁵⁸ On this pessimistic view, counterspeech, including state-based counterspeech, is wholly incapable of challenging the coded expression of vilifying contents.

But this conclusion is too strong. Having authoritative political figures speak back *is* sometimes a powerful way of blocking hateful coded presuppositions. Some of Stanley’s own examples suggest as much. For example, following Representative Paul Ryan’s comments asserting that “inner cities” lacked a “culture of work”, Representative Barbara Lee replied: “Let’s be clear, when Mr. Ryan says ‘inner city’ [...] these are simply code words for what he really means: ‘black’”.⁵⁹ While Ryan denied having said anything about race, the response prompted widespread criticisms in the media.⁶⁰

This last case suggests a more nuanced conclusion. Counterspeech can indeed block the dignitarian harms of coded degrading speech. However, the success of such counterspeech depends on having various skills and abilities, including: the ability to identify vilifying presuppositions; the clarity to challenge them in a way that makes it difficult for others to twist one’s words; and sufficient authority to be taken seriously when one publicly battles over words. Indeed, being an elected representative with a record of fighting for civil rights was likely part of what enabled Lee to be more successful than Gingrich’s interviewer.⁶¹

Nevertheless, because these conditions for success are highly demanding—even for public officials and their empowered agents—the objection from coded speech remains problematic, and constitutes an apt qualification of Brettschneider’s optimism regarding state counterspeech. Often, state-based counterspeech may fail to block hate speech’s dignitarian harms. Accordingly, my second and more fundamental reply does not deny this worry, but argues that it can be turned, with greater force, against the hate-banning society. Indeed, Waldron’s conception of hate speech bans is overly idealised if these are taken to infallibly block the dignitarian harms enacted by public hate speech.

The idealisation at hand is partly empirical. As Section 2 noted, Waldron’s argument sidesteps debates about the causal effects of hate speech. However, it does make

⁵⁷ Stanley (2015, 158).

⁵⁸ *Ibid.*, 161.

⁵⁹ *Ibid.*, 160.

⁶⁰ E.g., Blow (2014).

⁶¹ Chumley (2014).

empirical assumptions concerning the causal effects of bans: namely, that they are effective at eliminating hate speech, or at least at removing it from public arenas where it can undermine the assurance of dignity.⁶² Some commentators maintain that these assumptions are at best contentious, and at worst unfounded.⁶³

Even so, because the empirical evidence regarding the overall merits of bans remains mixed,⁶⁴ we should try to avoid relying on it. For our purposes, then, the main problem is the following. Advocates of bans must show that coercive legal bans are better at handling the cases that proved problematic for state counterspeech. But even in theory, bans seem singularly unpromising with respect to coded hate speech. Recall the coded sentence discussed above: ‘Inner cities’ reliance on food stamps is draining public resources’. In context, this presupposed *that blacks are lazy*. However, banning such speech seems extremely difficult. To begin, and most obviously, the epistemic problems which afflicted counterspeech apply here too: simply identifying instances of coded speech for the purposes of legislating or implementing bans can be difficult.

Moreover, two further problems arise specifically for bans. First, legal bans are blunt tools, whereas the very existence of coded hate speech shows that hate speech can easily shift in form. Hence, even if we identified and banned coded hate speech, hate speakers might easily adopt new codes to avoid the ban. The problem is not so acute with counterspeech. It is easier to speak out against a new instance of coded speech than it is to pass yet another law or even to implement a generally-worded law in a significantly new way. Notice that the problem here is moral, not empirical: what makes bans relatively inflexible is that, because they employ coercive force (whereas counterspeech does not), implementing them requires meeting a higher burden of moral justification. This (as Section 5 will discuss) is because coercive measures involve a pro tanto restriction of the coerced’s freedom. The point, then, is not just that the

⁶² If Waldron denied caring about the causal effects of bans on hate speech and asserted that he simply cared about their expressive power, this would generate two problems. First, it would contradict his earlier claim that he wants the public assurance of dignity to be implicit. Second, it would no longer be clear what legal bans offer which counterspeech does not, since counterspeech is centrally concerned with expressive powers. In fact, even if one believes that criminal law has *distinctive* expressive powers, as Green (2013) suggests, it does not follow that bans are necessary to play this expressive role. As many legal theorists observe, the expressive aim of denouncing the content of hate speech can be performed by many laws besides bans, such as anti-discrimination and civil rights laws. E.g., Post (1995, 26), Strossen (2012, 391), Heinze (2013, 595-596).

⁶³ For negative evidence, see Heinze (2016, 145-153) and Mchangama (2012).

⁶⁴ On these mixed results, see Gelber and McNamara (2015b).

institutional processes for implementing bans are in fact more inflexible than counterspeech—instead, it is that they are morally required to be.

Second, trying to ban coded hate speech exacerbates a common charge against hate speech bans, namely that they risk obstructing or ‘chilling’ democratic debates that are instrumental to promoting justice. Coded speech is widespread and works through otherwise innocent and useful terms, such as ‘welfare’, ‘food stamps’, ‘inner city’. Appealing to such terms is a normal and necessary part of fruitful democratic public deliberation. Consequently, *even if* we could identify and ban all instances of coded hate speech, this seems undesirable given how much banning this would involve, and how ordinary and useful the banned words or phrases would be.

Coded hate speech, then, also poses a problem, and arguably a greater one, for bans. Proponents of bans might perhaps respond that coded speech is less likely to arise in the hate-banning society. But this is implausible. First, it seems empirically dubious. In many European states that do have bans, extreme-right parties like the French *Front National* or the German *NPD* have gained mainstream prominence. And they have done so in part by smuggling hateful propositions into public deliberation under coded guises. In France, one of Marine Le Pen’s central strategies for mainstreaming the *Front National* has been to co-opt the language of republican values. It has been argued, for instance, that the French far-right has increasingly used words like *laïcité* (roughly: secularism) or ‘republicanism’ to refer specifically to opposition to Islam.⁶⁵ Equally, the *NPD* has avoided bans by substituting ‘Glory and Honour’ for the banned ‘Blood and Honour’, replacing the swastika with a Nazi-era triangular figure, and using a salute with three-outstretched fingers rather than an outstretched hand.⁶⁶ (Notice how this last example illustrates how easily citizens can evade bans by adjusting their communications.) At first blush, then, coded hate speech seems just as widespread in hate-banning societies.

This should hardly be surprising: there are theoretical reasons to expect coded speech to be *more* prevalent in societies that ban hate speech than in those that do not. First, the imposition of legal penalties incentivises the development of coded hate speech by citizens wishing to express abhorrent views. This is not to deny that some such incentives, like social opprobrium, exist in hate-allowing societies. But, in virtue of

⁶⁵ Mondon (2014); McDonnell (2015).

⁶⁶ Heinze (2016, 145-148).

operating through the use of coercive sanctions, legal bans produce *further* incentives. Because they do so, Eric Heinze argues, “bans [...] are teaching [hate groups] how to mock and scorn vulnerable targets from within the bounds of the law.”⁶⁷

Furthermore, the principle of operation underpinning bans not only incentivises, but also facilitates the development of coded hate speech. Underlying the view that we should ban hateful discourse, Waldron suggests, is the compelling thought that some propositions are simply not up for debate.⁶⁸ On this view, we need not discuss why, say, ‘Immigrants are vermin!’ is false. We need not explain that all human beings have equal dignity, that this dignity mandates treating people with respect, and so on. We simply ban the hateful utterance. And if some cannot understand why, so much the worse for them.

According to Robert Talisse, however, when politics is organised around such a principle of non-engagement, “the very content of the concepts that shape and direct our political institutions is left underdetermined; they are up for grabs [...] [and] available for easy co-optation”.⁶⁹ Here, Talisse is echoing John Stuart Mill’s warning that, in the absence of a vigorous defence of important truths, “the shell and husk only of [their] meaning is retained, the finer essence being lost”.⁷⁰ The implication, in our context, is that refusing to publicly repudiate hateful utterances ironically contributes to making the relevant values (‘dignity’, ‘respect’, ‘equality’, etc.) up for grabs. Other things being equal, it is easier for hate groups to present their views as egalitarian or concerned with dignity in contexts where no concrete articulation or reiteration of these values has recently been offered. For example, it might be easier for racial segregation to be publicly defended as egalitarian in circumstances where, because we prefer bans to counterspeech, there has been little public discussion of what equality truly means.

In contrast, democratic counterspeech actively employs powerful tools for authoritatively endowing democratic values with concrete content. As discussed in Chapter 2, it not only gives arguments for why hateful views are wrong, but employs emotionally-charged narratives to show precisely what is meant by democratic ideals. Think, for instance, of Martin Luther King’s ‘I Have a Dream’ speech, where he offers an impassioned vision of a society of equals, which explicitly aims to recapture “the true

⁶⁷ Heinze (2013, 600).

⁶⁸ Waldron (2012, 95-96).

⁶⁹ Talisse (2009, 64).

⁷⁰ Mill (2006[1859], 47).

meaning of equality”. Such a narrative contributes to authoritatively clarifying the meaning of equality in a way that makes co-opting this ideal more difficult.

Given these deficiencies with respect to coded speech, advocates of bans might be tempted to offer a different kind of response: that this section’s principal line of argument attacks a straw man, as they are actually not interested in banning coded hate speech.⁷¹ This response is inadequate for two reasons. Firstly, given the logic of Waldron’s argument, it is unclear that this response is genuinely available to him. After all, the rationale for banning public hate speech is that it undermines the public assurance of dignity and assures other hateful citizens that they are not alone. But this rationale does not seem inherently sensitive to whether speech takes a coded form or not. Indeed, as the *NPD*’s adjusted Nazi salutes illustrate, coded hate speech can be such that, while it is encoded enough to escape existing regulations, it clearly expresses and diffuses a dignity-denying message. Therefore, this response seems *ad hoc*.

Secondly, and more fundamentally, this reply is unsatisfactory given the dialectical situation. The discussion of coded speech was initially raised on behalf of critics of state counterspeech to suggest that such counterspeech is less effective than bans. My argument has been that coded speech is at least as problematic for bans as it is for counterspeech. In this context, then, advocates of bans cannot respond that they are not interested in regulating coded speech without also defusing the original objection to counterspeech.

There is another sense in which my argument might be accused of attacking a straw man. The defender of bans might say that they are not recommending legal bans *instead* of state-sponsored counterspeech. Rather, they want bans *in addition to* state-sponsored counterspeech. So, they can avail themselves of all of the comparative benefits of counterspeech.

This response, though natural, faces two difficulties. Firstly, it assumes that we can simply add the good effects of bans to those of state-based counterspeech by combining the two. But this assumption is contentious. We have encountered reasons for thinking that bans, because of their distinctive mode of operation, encourage and facilitate the development of coded hate speech. And the harms enacted by coded hate speech are especially difficult to block, for both counterspeech and bans. (This, note, is consistent with my argument that coded hate speech is even more problematic for bans

⁷¹ For comments suggesting this line of response, see Waldron (2012, 189-192).

than for counterspeech). State-sponsored counterspeech may therefore be less effective alongside bans than on its own.

To appreciate the second difficulty, consider once more the dialectical situation. In Section 3, I argued that bans involve a *pro tanto* moral cost, stemming from considerations of freedom (understood as non-domination), which state counterspeech avoids. To justify this moral cost, the advocate of bans needed to show that bans are, at least in some cases, more effective at blocking hate speech's harm than state-driven counterspeech. This section has suggested that this is not the case. But if bans offer nothing more than state-sponsored counterspeech by way of blocking the harm in hate speech, we still have a reason to prefer simple counterspeech to the combination of bans and counterspeech: the latter is no more effective than the former, yet involves a freedom-related cost that the former avoids.

I began by considering the objection that counterspeech is often ineffective at blocking hate speech's dignitarian harm, particularly when hate speech takes a coded form. While this objection is powerful, it does not establish that state counterspeech is entirely ineffective in these contexts. More importantly, this objection can be turned against bans. Not only do bans seem even more ineffective in response to coded speech, but it is inherent to their principle of operation that they encourage and facilitate the development of such speech.

5. Objection 2. Does Counterspeech Block Only Hate Speech?

5.1. The Risk of Misdirected Counterspeech

According to the previous objection, state-based counterspeech cannot block enough instances of hate speech. By contrast, the present objection claims that state-based counterspeech risks challenging *too many* instances of speech. That is, it risks undermining contributions to democratic public deliberation that are not hateful (for short, 'non-hateful speech') and that are even sometimes instrumental to advancing justice.

Robin West directs a variant of this objection at Brettschneider, by raising the possibility of the "Hypocritical State". So far, we have been assuming that the state will only use its powers of counterspeech against hateful speech. But part of what is non-ideal in the real world is that states and influential political figures sometimes use their

expressive powers wrongly. Put differently, they misdirect their speech. One explanation for why they do so, West observes, is that states are sometimes hypocritical. They appeal to values like dignity or equality to advance unjust ends that are contrary to those values, and to denounce speech that genuinely aims to promote those values.⁷² Therefore, recommending a policy of state counterspeech in non-ideal conditions risks undermining speech that is not hateful and that may instead be geared towards advancing justice.

The public reception of Black Lives Matter brings this concern to life. In 2016, the social movement campaigned regularly against structural racism directed at black Americans, and notably protested the violent treatment of blacks by law enforcement officers. In response, soon-to-be-president Donald Trump and Rudy Giuliani condemned the slogan ‘Black Lives Matter’ as racist on the fallacious grounds that it implied that non-black lives did not matter.⁷³ Here, West might want to say, we have a case of misdirected speech: authoritative public figures used the language of racial equality to denounce a public slogan that was neither hateful nor inegalitarian, and that instead promoted racial equality.

This is a serious objection. Again, however, it can be turned against the policy of banning hate speech. If, as West claims, the state is badly motivated, or hypocritical, then it might misdirect bans too. It might, say, appeal to the ideal of human dignity to ban speech that is not genuinely hateful, and that rather aims to advance human dignity.

There is some empirical evidence of bans being misused thus. Indeed, opponents of bans often complain that bans are routinely used to silence groups contesting genuine injustices.⁷⁴ That being said, I am not committed to the claim that bans are actually often misdirected. Rather, my principal claim is a conditional one: *if* there is reason to think that the state’s expressive powers will be misdirected, then, *prima facie*, there is reason to think that the state’s coercive powers too will be misdirected.⁷⁵ After all, if the state is ill-motivated, as West suggests, why think that this will not affect its usage of bans as well as its counterspeech? So, *insofar* as this objection is effective against the policy of state counterspeech, it should be effective against the policy of banning hate speech.

⁷² West (2014, 1039).

⁷³ Twohey (2016); Weigel (2016).

⁷⁴ E.g., Mchangama (2012, 99), Heinze (2013, 592).

⁷⁵ For a related concern which considers the Trump presidency, see Kahn (2017, 246-251).

Once again, then, the hate-banning and hate-allowing societies appear to be companions-in-guilt. Further, the hate-banning society is once more the guiltier party. Misdirected banning—misusing bans to penalise non-hateful speech—is more morally problematic than misdirected counterspeech—speaking back against non-hateful speech. First, banning involves coercive power, rather than merely persuasive or expressive power.⁷⁶ Since coercive power involves employing force to make someone do or not do something, it diminishes individual freedom or autonomy in a way that merely expressive or persuasive power does not.⁷⁷

Second, the coercive power involved in misdirected banning also reduces individual freedom because of *what* it keeps individuals from doing, namely publicly voicing considerations that matter to them. As Section 3 discussed, being free to publicly raise such considerations is valuable partly because it enables speakers to contest exercises of political power and thereby hold them accountable. Consequently, by punishing non-hateful public speech, misdirected banning risks significantly undercutting the freedom (understood as non-domination) of its targets.

Now, we should not overstate the moral asymmetry between misdirected banning and misdirected counterspeech. Although misdirected counterspeech does not employ coercive force, it too can impede its targets' freedom. To see this, recall that authoritative counterspeech *can* disable certain speech-acts. In Section 3, the basis for recommending state counterspeech was that it can disable the speech-act of refuting the public assurance of dignity. In the case of misdirected counterspeech, state counterspeech diminishes the target's perceived authority, and thereby reduces her capacity to *convince* her audience through non-hateful speech. Thus, insofar as being able to publicly raise concerns and convince one's listeners protects one against political domination, misdirected counterspeech also impedes freedom as non-domination.

Nevertheless, the crucial point remains that misdirected counterspeech does not disable *all* speech-acts associated with the targeted utterance. In voicing her non-hateful concerns, the speaker can still successfully *disclose* her worldview, *share* her experiences, and *expose* others to new arguments. Further, the speaker can *contest* the ill-motivated state's claim that her views are unjust by offering a rival understanding of the relevant values. For example, when leading politicians fallaciously condemn the slogan 'Black

⁷⁶ Brettschneider (2012, 6-7, 81).

⁷⁷ Raz (1986, 377-378).

Lives Matter’ as inherently racist, members of Black Lives Matter still can (and do) contest that claim: they can argue, say, that when one racial group faces special injustices, achieving racial equality requires paying special attention to that group’s interests. Such speech-acts at least help third parties better understand the speaker’s non-hateful concerns, so that some of her discursive power to advance her goals is preserved.⁷⁸ Hence, a morally-important asymmetry persists:⁷⁹ while misdirected counterspeech salvages some valuable speech-acts associated with the targeted utterance, misdirected banning suppresses them all.

5.2. Can Bans Avoid the Objection?

Section 5.1 argued that the problem of misdirected state action is more problematic for bans than it is for counterspeech. Consequently, proponents of bans might be tempted to deny its basic premise: namely, that the state is likely to be ill-motivated. Waldron voices such a reply. Put differently, he suggests that the objection under consideration—that if we allow some hate speech bans, the state will ban some things that it should not—is overblown. Like many slippery-slope arguments, it relies on unsubstantiated empirical claims, which are overly pessimistic about states’ motivations.⁸⁰

One line of response asserts that these claims *are* substantiated, and that states do tend to use bans improperly.⁸¹ To avoid relying on such empirical claims about states’ motivations, I will offer a different response, which focuses on the epistemic difficulties involved in applying hate speech regulations. Specifically, I will argue that there are conceptual reasons, grounded in the vagueness of the harm in hate speech, that explain why states are liable to misidentify instances of hate speech, and hence—if implementing bans—to ban non-hateful speech.

To see this, let us re-examine the harm in hate speech. Waldron is at pains, firstly, to distinguish the harm that warrants legal regulation from the notion of offense. This is because offense is too widespread. “Especially in a multifaith society, religion is an area

⁷⁸ Notice a difference with speaking back against *hateful* speech. State counterspeech reduces the persuasiveness of hate speech because it has authority, and because it has the force of better argument on its side. But with *misdirected* counterspeech the speaker against whom the state speaks may well offer more compelling reasons. So she retains some persuasive force in a way the hate speaker does not.

⁷⁹ Rowbottom (2017, 33-34).

⁸⁰ Waldron (2012, 202).

⁸¹ Mchangama (2012, 97-99).

where offense is always in the air. Each group's creed seems like an outrage to every other group".⁸² Consequently, religious freedom cannot be preserved without freedom to offend.⁸³ A second desideratum for a speech-induced harm which we can safely recommend banning, Waldron suggests, is that it must not be vague.⁸⁴ If it is vague—if it is often indeterminate whether the harm obtains or not—then there is a conceptual reason to expect that whoever is implementing the ban will make mistakes and ban non-hateful speech. The harm's vagueness would constitute an underlying cause for why legislators and judges will make mistakes.

Putting these two desiderata together, a minimal condition for the harm in hate speech to be one that we can safely recommend banning is that it be clearly, rather than vaguely, distinct from offense. For Waldron, the harm of assault on dignity, where dignity is the status of being an equal citizen in good standing, satisfies this condition. Being offended is a subjective or psychological harm, which manifests itself in felt experiences like shock or distress. By contrast, assault on dignity is an objective harm, which concerns "how things are with respect to [one] in society".⁸⁵ Whether one's dignity is assaulted depends fundamentally on whether speakers are actually rejecting one's good standing. Conceivably, there might be cases where one feels as though this is the case though it is not. Conversely, while shock and distress often accompany assaults on dignity, they are neither constitutive nor necessary consequences of such assaults. Thick-skinned targets of dignity-denying speech, for instance, may not feel distressed. Because of this qualitative difference, Waldron is confident that the distinction between offense (which should not be banned) and assaults on dignity (which should be) is clear, not vague.

Ultimately, however, this move fails to dissipate the vagueness of the harm in hate speech. Even if assaults on dignity are objective harms, and therefore of a different ontological kind from offense, the distinction remains *epistemically* vague. Put differently, even if they are two different kinds of things, the way in which we *find out* whether they are occurring often does not allow us to distinguish the two.⁸⁶ This is because,

⁸² Waldron (2012, 127).

⁸³ Ibid., 130.

⁸⁴ Ibid., 136-143. This may seem inconsistent with Waldron's insistence elsewhere that vague laws can be adequately action-guiding. But there is no inconsistency. When Waldron (2010, 75-77) considers the objection that vague laws might suppress too much behaviour, he replies that this may be a valid consideration with free speech, but not in the cases he discusses there (driving and torture).

⁸⁵ Waldron (2012, 106).

⁸⁶ On epistemic vagueness in ethics, see Dougherty (2014).

intuitively, the most natural way to find out whether an utterance assaults the assurance of other citizens' good standing is to consider how its targets experienced it. Did they feel a combination of, say, shock and outrage, shame and humiliation, which combined to produce an overall feeling of not belonging, of being excluded? And using these subjective epistemic indicators simply is looking at whether someone has been offended.⁸⁷

One might think that the subjective hurt resulting from dignitarian assaults is phenomenologically distinct from mere offense. But it is unclear why this would be. It seems perfectly conceivable that someone whose dignity has not really been assaulted could have a qualitatively similar psychological response as a target of dignity-denying speech. For instance, as Waldron himself notes, a believer confronted with a merely offensive image of Christ may well experience shock and outrage, shame and humiliation, and feel that being humiliated thus denigrates her social standing.⁸⁸

To avoid this problem, Waldron suggests using an alternative epistemic indicator. Instead of trying to distinguish assaults on dignity from offense based on the experiences they induce, the idea is that we should look more directly for objective markers of assaults on dignity.⁸⁹ These markers would help distinguish reasonable feelings of having one's standing attacked from unreasonable ones. What could these markers be? Waldron suggests that merely offensive speech denigrates a person's beliefs, whereas assaults on dignity denigrate the person herself.⁹⁰

This distinction, however, is difficult to sustain. Most problematically, it struggles to handle the case of religious offense, which Waldron was concerned with distinguishing from assaults on dignity. This is because religious believers might respond that their religious beliefs are constitutive of their identity, so that deriding those beliefs is also an attack on their persons. Consider again the 2005 Danish Cartoons, which represented the prophet Muhammad as a bomb-throwing terrorist, or *Charlie Hebdo's* representation of Muhammad as a child-molester. By attacking central figures of Islam, the response would go, these attack the dignity of profound adherents to Islam. So, this objective criterion fails to clearly rule out religious offense.

⁸⁷ For instance, Gelber and McNamara (2015a, 336-337) use self-reports of subjective experiences to determine whether individuals' dignity has been assaulted.

⁸⁸ Waldron (2012, 114-115).

⁸⁹ Ibid., 115.

⁹⁰ Ibid., 120.

Waldron considers a similar response but dismisses it as “largely an irresponsible attempt” by its advocates “to claim more [...] than they are entitled to” by “recklessly presen[ting] claims about offense as though they were non-negotiable”.⁹¹ And when confronted with the Danish cartoons, he grudgingly concludes that they merely constitute a bad offense: attacking Muhammad’s dignity is not attacking Muslims’ dignity.⁹²

This reply seems uncharitable. We can easily imagine a reasoned case for thinking that attacks on Muhammad also constitute attacks on Muslims. ‘Since part of what it is to be a Muslim is to hold Muhammad’s life up as an exemplar’, someone might say, ‘such representations entail that we Muslims *aspire* to be terrorists!’ Arguments like this suggest that, even if Waldron were right that many religious beliefs are not constitutive of believers’ identity, there is nothing clearly absurd or “irresponsible” in thinking so. In religious contexts, then, the distinction between attacks on persons and attacks on beliefs (and accordingly between reasonable and unreasonable claims to having one’s dignity assaulted) often remains epistemically vague.

Moreover, Waldron’s reply jars with one of the more appealing features of his framework: his methodological commitment to asking what it is like for those on the receiving end of potentially harmful speech to be in this position. By introducing criteria of reasonableness to weed out unreasonable feelings of exclusion, and dismissing interpretations of these criteria voiced by self-described targets of religious hate speech, Waldron seemingly departs from this commitment. The point is not just to indicate an internal inconsistency. As I argued earlier, this methodological commitment is attractive. It is sensitive to the fact that targets of disrespectful speech may have a privileged perspective on what may be harmful about it. Furthermore, even when targets are mistaken regarding its harmfulness, it helps us better understand how they experience the world and thus facilitates engaging with them.

This section has examined the objection that state counterspeech may be misdirected in non-ideal contexts. This objection, I have shown, applies even more forcefully against bans. This is so even if, like Waldron, one rejects the claim that states are badly motivated. Even if those in charge of banning hate speech are not badly motivated, the basis for banning hate speech remains epistemically vague. Hence, the

⁹¹ Ibid., 131.

⁹² Ibid., 126.

present objection stands, and continues to stand more strongly against bans. To make this more vivid, let us end with an example. When it is vague whether citizens' dignity has been assaulted—as for example with the Danish cartoons or *Charlie Hebdo* cartoons—the hate-allowing state may make mistakes regarding whether to speak back or not. Suppose it does not speak against *Charlie Hebdo*. Instead, it denounces those who publicly excoriated the cartoonists. And suppose this is a mistake. Even then, those against whom the state speaks will not, by contrast with the banning case, be entirely silenced. Accordingly, we can at least hope to achieve a better understanding of why some citizens experienced the cartoons as an assault on their dignity. And if some of the cartoons did constitute unjust dignitarian assaults, this may help us recognise this fact.

6. Conclusion

How should we respond to public narratives expressing deeply disrespectful or hateful views? Public hate speech assaults the basic political status of its targets. This harm, it has been argued, warrants implementing legal norms that suppress such speech: unless we ban hate speech, democratic public deliberation risks promoting injustice rather than justice. While this diagnosis of the harm in hate speech is compelling, the inference to hate speech bans is too direct. It relies on an idealised picture of bans and an unsophisticated understanding of democratic counterspeech. This chapter has aimed to remedy this shortcoming and thereby offer a qualified defence of the view that public hate speech should be countered via democratic public discourse, rather than bans.

On a refined understanding of the counterspeech approach, the state authoritatively speaks back against hate speech despite not banning it. Given the freedom-related costs involved in banning hate speech, especially in conditions of widespread political alienation, such democratic counterspeech seems a preferable way of promoting dignity and advancing towards greater justice. Nonetheless, advocating state counterspeech might seem naïve: in non-ideal conditions, it is likely to achieve both too little and too much. My strategy in responding to these objections has primarily been to argue that counterspeech and bans are companions-in-guilt. Bans too struggle to block the dignitarian assaults produced by coded hate speech. And bans seem just as liable to be

misdirected as state counterspeech. In fact, in both cases, these objections seem more problematic for bans than for counterspeech. Bans are the guiltier party.

However, this defence of robust norms of state-sponsored counterspeech over legal bans must be qualified. First, my defence of counterspeech remains pessimistic. The companions-in-guilt structure of my argument shows that democratic counterspeech is far from a perfect response to abhorrent public narratives. One of the unhappy conclusions of this chapter is that, in non-ideal conditions, both attempted solutions face serious limitations.

Second, as I noted earlier, I do not purport to have definitively shown that bans on public hate speech are never warranted. My argument against bans and for democratic counterspeech is defeasible for two reasons. To begin, there is enormous variation in the forms that abhorrent speech takes and in the contexts where it occurs. Inevitably, I have focused on a restricted set of cases—in particular, cases of public discourse occurring in relatively stable democracies. In other types of political contexts, the companions-in-guilt arguments I have supplied may be less applicable.

Moreover, I have been focusing on one important speech-based harm: how hate speech constitutes an unjust assault on the public assurance of dignity. But there are other speech-based harms. For some of the causal harms of hate speech, analogous companions-in-guilt arguments, which would rely on substantial empirical investigation, may perhaps be less successful. Further empirical research is therefore needed to ascertain how far the argumentative strategy I have deployed for defending counterspeech extends to other speech-based harms.

Still, an important upshot remains: insofar as we are concerned with upholding the assurance of citizens' equal standing in stable democracies in the face of public narratives expressing abhorrent views, there are good reasons for thinking that democratic counterspeech is a preferable solution to bans.

CHAPTER 4

RAGE INSIDE THE DELIBERATIVE SYSTEM

1. Introduction

1.1. Anger and Visibility

In the classic novel *Invisible Man* (1952), Ralph Ellison depicts the perspective of an unnamed black American whose skin colour renders him “invisible”: he is persistently ignored, misunderstood, or mistaken for another. In a final moment of introspection, the invisible narrator turns to the reader:

So why do I write, torturing myself to put it down? Because in spite of myself I've learned some things [...] Why should I be the one to dream this nightmare? Why should I be dedicated and set aside—yes, if not to at least *tell* a few people about it? There seems to be no escape. Here I've set out to throw my anger in the world's face [...] ‘Ah,’ I can hear you say, ‘He only wanted us to listen to him rave!’ But only partially true: Being invisible and without substance, as it were, what else could I do? What else but try to tell you what was really happening when your eyes were looking through? And it is this which really frightens me:

Who knows but that, on the lower frequencies, I speak for you?

The passage is fascinating, not least for what it suggests about the communication of anger. First, the narrator emphasises the importance of “throwing [his] anger” at his audience. Further, he seems to do so for epistemic reasons, to make visible what had previously been invisible, by foregrounding his nightmarish experiences. Relatedly, perhaps, he claims that in expressing his anger, he speaks for the audience.

Interestingly, the invisible narrator is not alone in voicing such thoughts. His suggestions resonate strongly with the poet Audre Lorde's assertion that “my anger and your attendant fears, perhaps, are spotlights that can be used for your growth, [...] for

corrective surgery”.¹ Or consider, similarly, the novelist and civil rights advocate James Baldwin, who once declared that “to be a Negro in [America] and to be relatively conscious, is to be in a rage almost all of the time”, before recommending “control[ling] the rage so that it won’t destroy you”, thereby turning it into a productive force. For him, as a novelist, doing so manifests itself in the process of “creat[ing] a person and mak[ing] other persons feel what this person feels”.² A pattern emerges.

In this chapter, I examine the place of anger in democratic public deliberation. In doing so, I will try to make sense of the recurrent themes above. How is publicly expressing anger “corrective surgery” for one’s audience? What is it capable of teaching us, or making us feel? And why does it matter that, as the invisible narrator suggests, the angry speaker be “dedicated and set aside” for this purpose? These recurrent observations, I will argue, point to important insights regarding anger’s function and place in democratic public discourse—insights, moreover, which extant philosophical discussions of anger have sometimes hinted at, but never precisely captured.

1.2. Outline

Consider how this fits within my broader inquiry. In Chapter 2, to counter the objection that the Shared Reasons Constraint excludes too many things from public debate, I recommended giving a greater role to emotionally-charged narrative in democratic discourse. But a powerful objection to this recommendation is that it risks being too *inclusive*. Absent further qualifications, it is liable to allow, for instance, a racist’s enraged narrative. There are two worries here: one concerns the content of the racist views being expressed; the other bears on the rage, or intense anger, the speaker expresses, independently of what that rage is about. Chapter 3 considered the first worry, by investigating how we should counter public expressions of deeply disrespectful views. The present chapter turns to the second worry.

To keep the two worries apart, I will focus on cases where the contents of the angrily communicated speech are not immoral or disrespectful. Imagine, for example, an American civil rights activist in the 1960s, who affirms the principle of racial equality and denounces as unjust the fact that blacks are denied equal rights. And suppose she

¹ Lorde (1997, 278).

² Baldwin et al (1961, 205).

expresses these views in a way that is charged with deep anger. In this context, we can ask: is it desirable, from the standpoint of justice, that she publicly expresses this anger in conditions marked by deep divisions and pervasive injustices?³ Or ought she, morally, to refrain from expressing her views in a way that is suffused with anger?

Democratic theorists who explicitly invite emotions into their normative conceptions of public deliberation often intimate that this includes negative emotions such as anger. Krause, for instance, cites with approval the anger that members of the ACT UP publicly directed at Americans' failure to take the AIDS crisis seriously.⁴ Similarly, Gutmann and Thompson praise the intense outrage that Carol Moseley Braun voiced in the US Senate, concerning the proposed renewal of a patent on the Confederate flag insignia.⁵ According to these theorists, then, deliberators concerned with advancing justice needn't necessarily refrain from publicly expressing their anger.

But anger has come under attack. Spurred in part by the rage and resentment that pervade contemporary politics,⁶ some philosophers—including, most influentially, Martha Nussbaum—have revitalised a historical tradition of scepticism towards anger. By contrast with the case of disrespectful or vilifying hate speech, critics of political anger are typically not suggesting that legal norms should coercively suppress angry public speech. But they nevertheless hold that publicly expressing anger is morally undesirable. One of the most influential arguments for this view, an argument which is commonly echoed in broader public discourse, asserts that expressing anger is counterproductive from the standpoint of justice. Although it is an aversive reaction to injustice, it is likely to exacerbate existing injustices. Thus, even if some emotions have an important role to play in public deliberation, deliberators should refrain from voicing deep anger. Justice would be better served, critics conclude, by expressions of forgiveness, love, or even meekness.⁷

Against this popular challenge, I will argue that anger does have a key role to play in democratic public deliberation, partly by articulating an underexplored epistemic value of publicly-expressed anger, and partly by challenging the broader framing of the

³ I will only consider cases where speakers have some control over their expression of anger. After all, if they cannot refrain from expressing anger, recommending that they do so asks for something impossible.

⁴ Krause (2008, 119). See also Young (1996, 124), Mansbridge (1999, 223), and Chakravarti (2014).

⁵ Gutmann and Thompson (1996, 135-138). However, they suggest that the angry oration may have been a non-deliberative means to deliberation, rather than an instance of deliberation.

⁶ Nussbaum (2016b).

⁷ On forgiveness, love, and meekness, see respectively McGary (1982), Nussbaum (2016a), and Pettigrove (2012).

debate. My argument proceeds as follows. After offering a working definition of ‘anger’ (Section 2), I provide an overview of the counterproductivity objection, and of the main strategies for addressing it. I suggest that even if we grant the non-consequentialist value of anger, we cannot circumvent the issue of whether anger is productive in terms of its consequences. Accordingly, I distinguish two complementary philosophical strategies for defending anger’s productivity, both of which I will pursue (Section 3).

The first part of my defence theorises an important positive consequence of political anger. Drawing on contemporary philosophy of emotion, I argue that public expressions of anger possess a distinctive epistemic value. In virtue of anger’s characteristic phenomenology, communicating anger to one’s listeners can play an indispensable role in helping them register and understand existing injustices. This epistemic role is crucial in divided societies, where the injustices suffered by some groups are often invisible to, or misunderstood by, others (Section 4).

On its own, however, this initial argument is insufficient to stave off the counterproductivity objection. To complement it, the second part of my defence recommends rethinking what constitutes the correct context for evaluating anger’s productivity. By appealing to democratic theory’s ‘systemic turn’, I contend that anger’s consequences should be evaluated not in isolation—as advocates and opponents of anger have often done—but rather from a systemic perspective. This reframing reveals that the worries concerning anger’s alleged counterproductivity are substantially more tractable than they initially appear (Section 5). Having defended the place of anger in the deliberative system, I conclude by revisiting Ellison’s invisible narrator (Section 6).

2. Defining Anger

To define our subject more precisely, let us outline some key properties of the emotion I am calling ‘anger’. Like all emotions, and unlike moods, anger is an intentional attitude, in the philosophical sense of the term—an attitude that is directed at a particular object, its content.⁸ Anger, put differently, is about something: A is angry *that B betrayed him*; C is angry *that she was not promoted*.

⁸ On emotions’ intentionality, see Deonna and Teroni (2012, 3-6), Brun et al (2009, 11-12), and Nussbaum (2015, 42-43).

Secondly, anger is a cognitive attitude: an important part of what anger does is aim accurately to represent certain features of the world. More specifically, philosophers commonly suggest that anger presents the content it is directed at as involving a moral violation or injustice. Because anger has this cognitive dimension, we can assess it as being more or less correctly directed, or *fitting*. Anger is fitting, firstly, only if its content actually involves an injustice; and, secondly, only if its intensity is proportionate to the severity of the injustice it is purporting to represent. Rage is not a fitting response to a minor moral violation, such as your borrowing my favourite pen without permission.⁹

Thirdly, anger is phenomenologically distinctive: there is a distinctive qualitative experience that is involved in being angry at something. This experience is partly characterised by certain bodily feelings, such as an increased heart rate, feeling hot, or trembling. However, anger's distinctive phenomenology also concerns the content it is directed at. A common observation among philosophers of emotion is that although emotions are directed at objects that are already given to us through perception, imagination, memory, or belief, emotions do not leave the representation of these objects unchanged. Instead, they are "distinctive ways of seeing a situation".¹⁰ In Section 4, I will say more about how anger characteristically affects the way we experience its content, and why this matters politically.

Finally, besides having a cognitive dimension, anger has a conative aspect. Not only is anger an attitude which aims to represent something about the world, but it also disposes us to action. For example, A's rage at the unfair treatment of blacks in the criminal justice system may motivate A to retaliate against law enforcement agencies, or to protest for their reform.

With this picture of anger in place,¹¹ we can define this chapter's focus more precisely. As we will see shortly, Nussbaum has mounted an influential challenge to anger, which centres on its conative dimension: anger is problematic, she claims, because it disposes us to retributive or vengeful actions, actions aimed at inflicting

⁹ On the fittingness conditions of anger, see Cogley (2014, 201-203), Chakravarti (2014, 3-6), Srinivasan (2017, 6-7), and Nussbaum (2015, 42-43).

¹⁰ Jones (1996, 11).

¹¹ This picture of anger is compatible with many theories of the emotions, including the perceptual theory (e.g., Döring (2007), Tappolet (2011)), and the attitudinal theory (e.g., Deonna and Teroni (2012, ch. 7)). And although, as I discuss later, it is unclear whether the evaluative judgment theory (e.g., Nussbaum 2015) can successfully explain anger's distinctive phenomenology, advocates of this theory typically *acknowledge* that emotions are phenomenologically distinctive. For discussion, see note 64, this chapter.

payback on the perpetrators of injustice.¹² But even Nussbaum allows for *some* forms of anger, which she labels “transitional anger”, and of which Martin Luther King is the paradigmatic illustration.

While transitional anger’s most distinctive attribute is that it is not retributive,¹³ Nussbaum’s discussion of King makes it clear that it has further distinguishing properties. A second characteristic is that although, like anger generally, transitional anger is directed at a morally deficient object, it swiftly transforms into a forward-looking attitude, which is oriented towards advancing justice. Thus, someone experiencing and expressing transitional anger does not dwell on past injustices.¹⁴ Thirdly, although transitional anger sometimes moves individuals to non-retributive violence (for instance, in self-defence), it habitually motivates non-violent actions.¹⁵ Finally, transitional anger has a distinctive tone. Its typical expression is not fiery or harsh, but calm and self-controlled.¹⁶

In this light, to avoid arguing against a straw man when defending anger’s place in public discourse, I will focus on non-transitional forms of anger. To this end, this chapter focuses predominantly on the public speech of Frederick Douglass (1818-1895), an abolitionist leader who escaped from slavery, and Malcolm X (1925-1965), who famously militated against the oppression of black Americans. Both were heavily involved in the struggle for racial justice, and both were reputed for their intensely angry rhetoric.

Crucially, the anger they expressed was seldom transitional anger. First, its conative dimension, or aim, could be retributive. Douglass, for instance, once asserted that slaveholders “deserve to have [their throats] cut”.¹⁷ And even when its aim was not retributive, their anger nonetheless typically differed from transitional anger. It was often agitated and harsh, sometimes shading into rage, it did not shy away from encouraging violence, and it often dwelled on past and current injustices. Consider, for example, Malcolm X’s insistent depictions of the “nightmare” constituted by black

¹² Nussbaum (2015, 45-48).

¹³ Ibid., 44. See also Nussbaum (2016a, 212)

¹⁴ Nussbaum (2015, 52-54).

¹⁵ Ibid. See also Nussbaum (2016a, 212, 221).

¹⁶ Nussbaum (2015, 52-54; 2016a, 222, 228-230).

¹⁷ Cited in Oakes (2007, 100).

Americans' daily lives—a rhetoric which contrasted so starkly with King's own forward-looking rhetoric of “dreams” that King repeatedly distanced himself from it.¹⁸

Besides these dialectical reasons, there is also a political reason for examining the value of non-transitional anger—namely, that such anger is widespread in non-ideal public discourse. Douglass and Malcolm X, with their deeply angry rhetoric, played highly influential roles in their respective struggles. What is more, the intensity of their anger echoes that of the three figures we started with. Baldwin, in particular, makes it clear that it is *rage* he needs to control and channel. So, while political philosophers have given due attention to the rhetoric of more moderate speakers, like Martin Luther King,¹⁹ a normative assessment of democratic deliberation in non-ideal conditions should pay greater attention to the more controversial, but nonetheless politically salient, case of non-transitional anger.

3. The Counterproductivity Objection: An Overview

3.1. The Counterproductivity Objection

One of the most common charges against publicly expressing non-transitional anger is that doing so is counterproductive: it is likely to amplify rather than alleviate existing injustices. This accusation pervades actual political discourse. Social movements that angrily denounce, say, police brutality against blacks, discrimination against women, or abuse directed at the LGBT community are frequently told that their advocacy would be more effective if it were less shrill.²⁰ But this concern also has a distinguished philosophical history. Perhaps most famously, Seneca asserts that anger is inconsistent with humankind's ethical ends: “mankind is born for mutual assistance; anger for mutual ruin”.²¹ Thus, “anger has nothing useful in itself”.²² The view that anger and its expression are counterproductive has also been revitalised by contemporary philosophers. Glen Pettigrove, for instance, suggests that because anger may interfere with interpersonal communication, blind angry individuals, and motivate oppressive

¹⁸ Baldwin (1986).

¹⁹ E.g., Dryzek (2010), Rawls (1997).

²⁰ For examples, see Srinivasan (2017, 1-5).

²¹ Seneca (1928[c. 45], I.5).

²² Ibid., I.9. See also Adam Smith (1982[1759], I.ii.ch. 4).

actions, “the person who does not grow angry [...] will be better positioned to focus on promoting common goods”.²³

Among contemporary theorists, however, the counterproductivity objection owes its most prominent restatement to Martha Nussbaum. In *Anger and Forgiveness*, Nussbaum contends that, because of its tendency to motivate payback, “anger is always normatively problematic, whether in the personal or in the public realm”.²⁴ As we will see in the following section, Nussbaum is partly concerned with anger’s intrinsic moral standing. But when explaining what makes anger problematic in the public realm, she frequently emphasises its bad consequences for justice.²⁵

Firstly, Nussbaum claims, anger tends to be an irrational waste of energy, which distracts from the pursuit of justice. Anger typically motivates individuals to seek payback or vengeance. But seeking payback, Nussbaum claims, will rarely undo the wrong which prompted the anger. Hence, “if we ponder the futility of the payback wish [...] we quickly discover that non-anger and a generous disposition are far more useful”.²⁶

Secondly, Nussbaum continues, anger is likely to positively obstruct the pursuit of justice. By “breeding mistrust” and “increas[ing] the anxiety and self-defensiveness” of its targets, expressing anger undermines interpersonal cooperation.²⁷ And interpersonal cooperation seems necessary for the wide-scale promotion of justice. In the worst cases, by inciting payback and eroding trust, anger risks triggering a “cycle of blood vengeance”.²⁸ This is no doubt what King also feared when condemning Malcolm X’s violently angry rhetoric, which he claimed could “reap nothing but grief”.²⁹ Thus, expressing anger seems deeply misguided if we truly care about advancing justice. In the pursuit of justice, “anger is politically futile”.³⁰

Anger’s alleged counterproductivity has normative upshots for both formal and informal politics. For Nussbaum, formal political institutions ought to be governed by a spirit of non-anger.³¹ To this end, citizens who populate and uphold these institutions

²³ Pettigrove (2012, 347, 369-370). See also Silvermint (2017, 10-15).

²⁴ Nussbaum (2016a, 5).

²⁵ Ibid., 220.

²⁶ Ibid., 228.

²⁷ Ibid., 230.

²⁸ Ibid., 1, 230-233.

²⁹ Cited in Baldwin (1986, 403).

³⁰ Nussbaum (2016a, 236).

³¹ Ibid., 172.

should exercise emotional self-restraint and resist “the siren song of anger”.³² What does this mean for specific political institutions? In the judicial realm, Nussbaum suggests, court proceedings should not institutionalise victim impact statements, where wronged parties voice their grievances.³³ In the realm of transitional justice, Truth and Reconciliation Commissions (TRCs) ought to avoid providing an outlet for non-transitional anger. Instead, they should give expression to “generous forward-looking thoughts”.³⁴ Accordingly, Nussbaum commends Desmond Tutu insofar as he discouraged narratives of anger during the South African TRC.³⁵ And although Nussbaum never discusses formal legislative politics, her condemnation of public anger presumably applies there too: if publicly-expressed anger truly impedes justice, this seems just as problematic when making laws as when interpreting them.

Nussbaum likewise recommends that informal political contestation be governed by norms of non-anger. Indeed, she categorically praises Martin Luther King, Mohandas Gandhi, and Nelson Mandela, who “all repudiated anger as a matter of both theory and practice”.³⁶ According to Nussbaum, their concerted efforts to refrain from expressing deep anger and to discipline away the anger that they sometimes felt were crucial to the success of their social movements.³⁷ Because of this, Nussbaum asserts that we ought, morally, to emulate them: “the revolutionary non-anger of Mohandas Gandhi and Martin Luther King, Jr., is proposed not as a distant hope but as an immediate task, to be embraced here and now in the confrontation with justice”.³⁸ Thus, participants in informal political contestation who wish to advance justice should refrain from expressing non-transitional anger.

³² Ibid., 197.

³³ Ibid., 194-195. Nussbaum (2016a, 196-197) does cautiously consider allowing such statements *post-sentencing*, but precisely because this would diminish their political impact.

³⁴ Nussbaum (2016a, 240).

³⁵ Ibid. However, Nussbaum goes further than Tutu: because she worries that even forgiveness might be a covert form of anger, she recommends that TRCs advance a narrative of unconditional generosity. For discussion of how the South African TRC discouraged expressions of anger, see Chakravarti (2014, ch. 2).

³⁶ Nussbaum (2016a, 212).

³⁷ Ibid., 227, 231-233, 236. Nussbaum (2016a, 236) acknowledges that King and Mandela occasionally experienced and even expressed anger. But she insists that, crucially, this anger was fleeting: it was (and needed to be) swiftly followed by a transition to forward-looking love and compassion.

³⁸ Nussbaum (2016a, 218).

3.2. Consequentialist and Non-consequentialist Strategies

A direct strategy for responding to this objection consists in arguing that public anger actually needn't be counterproductive: often, it may have more good effects than bad effects. However, to sidestep the thorny issue of whether or not anger is counterproductive, Amia Srinivasan suggests an alternative strategy. According to Srinivasan, the question of whether or not political anger is desirable is not fully answered by an assessment of anger's consequences. Even if we ascertain that overall, we have instrumental reasons not to get angry, there may still be intrinsic reasons to do so, reasons unrelated to anger's consequences. "For any instance of counterproductive anger we might still ask: is it the fitting response to the way the world is? Is the anger, however unproductive, nonetheless *apt*?"³⁹

What makes anger apt? Recall that anger has a cognitive dimension: it is directed at an object, which it represents as involving a moral violation or injustice. Anger is cognitively fitting only if its object actually involves an injustice and only if its intensity is proportionate to the severity of that injustice. For Srinivasan, aptness is the intrinsic moral value that accrues to anger in virtue of being a fitting cognitive response to injustice. The idea is that, just as there is intrinsic aesthetic value in appreciating beauty, so too there is intrinsic moral value in anger that correctly registers injustice.⁴⁰

Because anger is apt in virtue of fittingly or correctly registering its object's moral properties, its aptness does not depend on producing good consequences. Therefore, even if anger is instrumentally problematic, its intrinsic value suggests that we might yet be able to resist the conclusion that it is all-things-considered morally undesirable.

One worry with this response is that, for Nussbaum, anger's tendency to motivate retributive action, or payback, is not merely instrumentally problematic. It also makes anger questionable from a non-consequentialist perspective. On Nussbaum's view, either retribution is driven by the thought that inflicting suffering will somehow undo the injury which prompted one's anger, or retribution inflicts suffering in order to degrade or "down-rank" the offender's status.⁴¹ In the latter case, retribution seems immoral, displaying an obsession with rank.⁴² In the former case, retribution is

³⁹ Srinivasan (2017, 4).

⁴⁰ Ibid., 5-9.

⁴¹ Nussbaum (2015, 41-51).

⁴² Ibid., 51.

irrational, since only “magical thinking” could make one believe that inflicting suffering will undo the injury one has sustained.⁴³ Therefore, either the desire for payback involves an immoral obsession with the relative status of others, or it displays an irrational disposition to produce futile suffering. Since both traits are vicious, the fact that anger’s conative dimension inherently disposes us to retributive action means, Nussbaum concludes, that anger is intrinsically immoral.

Srinivasan, however, gives us reasons to doubt this. She denies that such a tight connection obtains between anger and retribution. “Anger without the desire for revenge is something many of us know well”. When a friend betrays me, Srinivasan suggests, it is perfectly conceivable that my ensuing anger might motivate me simply to seek his recognition that he has wronged me. This desire for recognition, it seems, has no essential connection to the desire for revenge.⁴⁴ If this is right, the upshot is that anger is not inherently associated with payback. Though anger may sometimes cause revenge, it is not itself essentially vindictive. So, even accepting for the sake of argument that retribution is inherently immoral, it would not follow that anger is intrinsically immoral.

But even if Srinivasan’s response is right (as I believe it is), her non-consequentialist strategy nevertheless does not allow us to circumvent the counterproductivity objection altogether. As Srinivasan herself recognises, the non-consequentialist reasons for expressing anger do not have absolute force. In other words, the aptness of anger gives us a *pro tanto* moral reason in favour of publicly expressing apt anger, but this moral reason could in principle be overridden by countervailing moral reasons.

The counterproductivity objection purports to give such countervailing moral reasons: if publicly expressing anger is indeed counterproductive, this gives us *pro tanto* moral reasons not to do so. And if expressing anger is *highly* counterproductive—if it is bound, as Nussbaum suggests, to impede crucial justice-promoting reforms—then these consequentialist reasons against expressing anger may well outweigh the non-consequentialist reasons Srinivasan invokes. To illustrate: if expressing rage really would have caused Abraham Lincoln to fail in his attempt to abolish slavery, and Martin Luther King to fail in his leadership of the Civil Rights Movement, then one might plausibly think that, overall, caring about justice required that they refrain from

⁴³ Ibid.

⁴⁴ Srinivasan (2017, 7-8). In fact, some of the evidence Nussbaum cites suggests that anger commonly seeks control or justice, which are conceptually distinct from vengeance. E.g., Tavis (1982, 153-153).

expressing anger. Correspondingly, if expressing anger really is extremely counterproductive, it might seem that enraged orators like Douglass and Malcolm X were overall wrong, from the perspective of justice, to express themselves as they did. Hence, although I agree with Srinivasan's claim that anger has non-consequentialist value, we should not stop there. To strengthen the case for publicly-expressed anger, the rest of this chapter aims to tackle the counterproductivity objection more directly.

3.3. Two Types of Consequentialist Strategies

One might worry that directly resisting the counterproductivity objection's consequentialist claim is essentially an empirical task, which should devolve to social scientists. But philosophers can also play a part. Firstly, consider that when social scientists investigate the empirical consequences of public anger, they do not look randomly, but instead proceed with some theoretical hypotheses regarding what types of effects might reliably be related to anger. The task of developing such hypotheses is one that philosophers can fruitfully contribute to. By elucidating emotions' phenomenology, intentionality, and relation to epistemic values, philosophical accounts of emotions help illuminate what *kinds of consequences* we can expect emotions to produce. Accordingly, one philosophical strategy for tackling the counterproductivity objection consists in identifying and conceptualising possible positive consequences of anger.

A second way in which philosophers can contribute is by examining what the correct *context* for evaluating the consequences of angry speech is. The context of evaluation is an account of what alternative things are being compared along a given standard of evaluation. Suppose that our aim or standard is to motivate others to oppose injustice, and that we have shown that publicly-expressed anger sometimes motivates listeners to oppose injustice. A critic of anger might acknowledge this, but reply that the relevant question is not: does angry speech motivate individuals to oppose injustice better than no speech at all? Rather, it should be: does angry speech motivate individuals to oppose injustice better than the best form of non-angry speech? Importantly, the criticism is not that angry speech produces a certain kind of (bad) consequence. Instead, it is that advocates of anger are taking the wrong objects as their comparanda. In this way, examining the context of evaluation constitutes a distinct

strategy through which philosophers can challenge claims about the productivity or counterproductivity of public anger.

I will pursue both strategies in tandem. Section 4 articulates a vital and neglected epistemic consequence associated with publicly-expressed anger. In turn, Section 5 contends that both advocates and opponents of the counterproductivity objection typically adopt the wrong context for evaluating public anger's consequences. Now, in doing so, my aim is of course not to show that publicly expressing anger will always be productive, all things considered. But it *is* to indicate that the counterproductivity challenge rests on at least two illicit moves: first, it overlooks a distinctive and important epistemic value of communicating anger; second, by adopting the wrong context of evaluation, it forces us to choose between these epistemic benefits and other values when doing so may be unnecessary. Establishing these conclusions will put us in a far better position to resist Nussbaum's charge that "anger is always normatively problematic,"⁴⁵ along with her recommendation that formal and informal political institutions be governed by norms of non-anger.

4. The Epistemic Value of Publicly-Expressed Anger

Opponents of the counterproductivity objection have enumerated various goods that anger can produce. Anger, it is said, motivates others to oppose injustice,⁴⁶ helps one retain one's self-respect,⁴⁷ and enhances one's perceived status.⁴⁸ Without rejecting these suggestions, I will specifically investigate the *epistemic* value of communicating one's anger. In part, this is because this dimension is particularly relevant to my broader inquiry: at various junctures, I have emphasised how democratic public deliberation matters, *inter alia*, because it contributes to pooling information that is relevant to just policy-making.⁴⁹

The view that anger and its public expression are epistemically valuable is not uncommon. Recall, for example, the invisible narrator's assertion that his anger aims to make visible what had been overlooked, and Lorde's reference to her anger as a

⁴⁵ Nussbaum (2016a, 5).

⁴⁶ Lorde (1997, 280).

⁴⁷ Bell (2009, 168); Borgwald (2012).

⁴⁸ Tiedens (2001).

⁴⁹ Chapter 1, Section 3; Chapter 2, Section 4.

“spotlight”. Furthermore, this view has been taken up by philosophers, particularly feminist philosophers, as part of a broader trend towards rehabilitating emotions within epistemology.⁵⁰

Nevertheless, the epistemic value of publicly-expressed anger has yet to be articulated adequately. To begin, some philosophers who discuss anger’s epistemic worth are not primarily concerned with the value of *publicly communicating* anger.⁵¹ And most of the defences that are concerned with publicly communicating anger are largely programmatic, rather than precisely developed.⁵² An exception here is Sonali Chakravarti, who has extensively explored the value of anger in victim testimony. But Chakravarti does not avail herself of recent developments in the philosophy of emotion, which—I will argue—are needed to understand what is distinctive about publicly-expressed anger’s epistemic contribution.⁵³ These limitations have enabled critics of publicly-expressed anger to acknowledge that it has some epistemic value, while downplaying the relative importance of that value. To pre-empt such replies, I will explore the epistemic value of publicly-expressed anger in a way that makes it clearer why this epistemic role matters, and why it cannot easily be performed without anger.

4.1. The Theoretical Case

Some epistemic benefits of publicly expressing anger are relatively uncontroversial. When public speakers express their demands with sincere anger—either by displaying the conventional physiological and vocal signs of anger, or by explicitly affirming that they are angry—we learn *that they are angry*. Given an understanding of what anger is, this reveals many things about their psychology. We learn that they are experiencing certain bodily feelings, like tenseness, that they have the sense that something is seriously unjust, and that they might be driven to action.

Although providing this information is an important epistemic function, two worries arise. Firstly, one might reply that this function could largely be performed without expressing full-blown anger. Nussbaum, for instance, might again point to

⁵⁰ E.g., Frye (1983), Jaggar (1989), Bell (2009), Chakravarti (2014, ch. 5). On this epistemological trend, see Brun et al (2008).

⁵¹ E.g., Jaggar (1989), Bell (2009).

⁵² E.g., Lorde (1997), Frye (1983).

⁵³ Chakravarti (2014, 2).

transitional anger. The entire cognitive content of transitional anger is “How outrageous. Something should be done about that.”⁵⁴ So, expressing transitional anger, as King did, suffices to communicate the belief that a situation is unjust, and the disposition to act on this basis. Accordingly, although Nussbaum acknowledges that non-transitional anger has some epistemic value as a signal of outrage and injustice, she readily downplays this value as “limited”.⁵⁵ Similarly, Pettigrove remarks that refraining from anger did not prevent Lincoln from expressing his moral opposition to slavery.⁵⁶ For example, in a letter to a political opponent, Abraham Lincoln declares: “You think slavery is right and ought to be extended; we think it is wrong and ought to be restricted. For this, neither has any just occasion to be angry with the other”.⁵⁷

Secondly, remember that we set out trying to illuminate the intriguing observations Ellison’s invisible narrator made about anger, observations which Lorde and Baldwin echoed. But the above account of publicly-expressed anger’s epistemic value does not fully capture what they meant. The invisible narrator declares that his anger teaches us something about the world, shining a light on nightmarish features we had overlooked. So far, however, the epistemic function of anger is merely that it teaches us something about the speaker’s psychology, not about the world she is looking at. Nor, moreover, does the present account elucidate Baldwin’s suggestion that useful rage works by making listeners feel what the speaker feels.

To isolate an epistemic function of publicly-expressed anger that can overcome these concerns, let us take a step back. Chapter 2 discussed how, in non-ideal conditions marked by deep divisions, logical arguments from shared premises sometimes break down. Often, the reason they do so is that some social groups lack certain concepts, or have erroneous views about the application conditions of certain concepts. This, in turn, typically results from the fact that they lack relevant experiences of the world. In this context, I argued, emotionally-charged narratives can remedy such communicative breakdowns, by facilitating the communication of one’s experiences and perspective to others. Applied to anger, the idea is that anger-infused discourse is epistemically important not just because it tells the audience *that the speaker is angry*, but also because it can help them imaginatively experience *what it is like to be in the speaker’s*

⁵⁴ Nussbaum (2016a, 6).

⁵⁵ Ibid., 211-212.

⁵⁶ Pettigrove (2012, 355).

⁵⁷ Lincoln (1860, 152).

shoes, how the world feels or appears from where they stand. Put differently, anger-infused narratives can enable the audience to empathise more fully with the speaker.⁵⁸

How does this work? The first thing to notice is that publicly expressing anger can be a means of transmitting anger to one's audience. Suppose A is describing an unjust state of affairs—say, slavery—to denounce it. But instead of simply enumerating facts about slavery, A expresses his intense anger, with the aim of arousing similar emotions in the audience. In doing so, A is making use of the propensity for human beings to resonate with, or 'catch' the emotions of others. As Chapter 2 mentioned, Hume famously observes this general tendency: in human beings, he suggests, "all the affections readily pass from one person to another".⁵⁹

This phenomenon is well documented in contemporary psychology, under the heading of 'emotional contagion'. Human beings tend to mimic the expressions, vocalisations, and postures of others, in a way that leads them to experience the emotions of others. This process is typically automatic, involuntary, and largely unconscious.⁶⁰ Importantly, this effect also holds with angry speech, specifically. One study, for instance, finds that when speakers raise their voices and accelerate their speech, "this is likely to raise the listener's blood pressure and feelings of anger".⁶¹ In short—and here we have an echo of Baldwin's thoughts about rage—given the tendency for emotional contagion, infusing one's narrative with anger can cause one's audience to feel that anger.

To illustrate, consider Douglass's indictment of slavery. Douglass, one historian records, was known for his fiery passion: "when he spoke, he roared, his booming baritone complemented by waving arms". And the fire was infectious: Douglass deliberately used this anger to "to make people feel—viscerally—the bloody horrors of slavery", "to provoke his listeners [...] to furious outrage".⁶² Nowhere was this effect

⁵⁸ For the identification of empathy and imaginative perspective-taking, see Matravers (2017, 1-2).

⁵⁹ Hume (2009[1738], 3.3.1.7).

⁶⁰ For discussion, see Barsade (2002) and Coplan (2011). Note that there can be other, more conscious, ways of coming to share others' emotions. Sometimes, Goldman (2011) suggests, we first imagine someone's perspective, then make an evaluative judgment on this basis, and this contributes to generating a relevant emotion. But I focus on emotional contagion for two reasons. First, as Coplan suggests (2011, 8), it is more reliable as a mode of emotional transmission. Second, the fact that the contagious transmission of emotion is largely unconscious is important for my defence of anger: it means that individuals can catch others' anger *prior* to judging that an injustice is involved in the relevant situation. This matters because I will argue that transmitting anger can help one's listeners register injustices that they had previously overlooked.

⁶¹ Siegman et al (1990, 641). For tentative evidence of contagion from angry facial expressions, see Blairy et al (1999, 35), Hess and Blairy (2001, 138-139), Friedman and Riggio (1981, 102-103).

⁶² Oakes (2007, 90-93).

more clearly reported than in William Lloyd Garrison's account of Douglass's first public speech:

I shall never forget [...] the *extraordinary emotion it excited* in my own mind—the *powerful impression* it created upon a crowded auditory [...] I think I *never hated slavery so intensely* as at that moment; certainly *my perception of the enormous outrage* which is inflicted by it [...] was rendered far more clear than ever.⁶³

What the passage eloquently reveals is the contagious nature of Douglass's anger, which spreads not just to Garrison, but to the audience generally.

Publicly expressing anger about x , then, can help one's audience experience anger about x . But why does this matter *epistemically*? As I noted earlier, philosophers of emotion commonly argue that emotional deliverances have a distinctive phenomenology, or felt quality. They are distinctive ways of seeing the object that they are directed at.⁶⁴ Hence, conveying one's anger to one's listeners helps them imaginatively undergo a distinctive qualitative experience. More specifically, the audience's perspective-taking, or empathy, is rendered more complete by emotional contagion. As the audience tries to adopt speaker A's perspective, their imaginative reconstruction is altered by the anger they have infectiously gained from A. Not only is Douglass's audience imagining the facts about slavery he is reporting, but the felt quality of their imaginings is coloured by the anger he has transmitted to them.

For our purposes, the fact that the audience's imaginative reconstruction of the speaker's perspective is coloured by anger matters only if this 'colouring' is epistemically valuable—that is, only if the phenomenology of anger enriches our understanding of the anger's object. To appreciate how it does so, let us consider two

⁶³ Garrison (1845, iv), emphases added.

⁶⁴ Nussbaum and Pettigrove dismiss the distinctive epistemic value of anger partly because they overlook this phenomenology. It is widely held that the distinctive phenomenology of emotion is irreducible to the phenomenology of a judgment, or even of a judgment/desire pairing. E.g., Deonna and Teroni (2012, 66-71), Goldie (2002, 73-74). But both Nussbaum (2015, 42-45) and Pettigrove (2012, 357-358) broadly take anger about x to be reducible to a judgment/desire pairing, involving roughly a judgment that x involves some moral violation, and a desire to retaliate against its perpetrators. Now, Nussbaum does try, elsewhere (2001, 65), to accommodate the rich phenomenology of emotions within her judgment theory of emotions. However, Ben Ze-Ev (2004, 454-457) and Deonna and Teroni (2012, ch. 5) forcefully rebut her attempt. Further, even if her account could accommodate emotional experience's complexity, the problem remains that Nussbaum's brief discussion of anger's epistemic value does not integrate these insights. She simply stresses that anger signals the judgment that something is morally amiss and that something should be done about it (2015, 55-56).

aspects of what philosophers of emotion have argued characterises the phenomenology of emotional representation.

The first relevant feature is that emotions are sources of salience: they draw our attention to certain aspects of a situation, and thereby place some properties of that situation into the foreground.⁶⁵ In this, we have a way to understand Lorde's reference to anger as a "spotlight". Emotions help us navigate complex environments by selecting features of the environment and highlighting them. In doing so, they "rende[r] previously ignored features and previously unknown patterns salient".⁶⁶

Which features do emotions call our attention to? Recall that the cognitive fittingness of an emotion depends on whether its object involves certain evaluative properties: loss for grief, danger for fear, injustice for anger, etc. Hence, one intuitive proposal is that a given emotion puts a spotlight on aspects of its object that are liable to ground those evaluative properties. Take the case of fear. If I am walking home at night and am afraid, my fear highlights features of the environment that may make it dangerous: the street's emptiness, the absence of lighting, and the footsteps behind me.⁶⁷ By extension, we should expect anger to highlight patterns of behaviour that are liable to ground injustices.

Putting these observations together, we have a first sense in which anger's phenomenology is epistemically valuable. Anger renders salient properties of our situation that are liable to ground injustices. Because the environments we navigate are extremely complex, we may otherwise fail to notice these properties. Thus, perceiving a situation with anger can make us notice an injustice that we would otherwise have overlooked. Furthermore, having our attention directed to the properties which ground a given injustice can help us see *why* something is unjust. Hence, the salience role of anger is epistemically valuable not only because it can yield knowledge that an injustice is occurring, but also because it can facilitate a greater understanding of the nature of that injustice.

A second characteristic of emotions' phenomenology suggests that emotions are epistemically valuable: in virtue of having a quasi-perceptual felt quality, it is commonly argued, emotional representations of objects allow us to register evaluative properties in

⁶⁵ On the salience role, see Elgin (2008, 43-46), Deonna and Teroni (2012, 122), and Brady (2013, 16-25).

⁶⁶ Elgin (2008, 45).

⁶⁷ Ibid., 43-44.

a way that can be more fine-grained than our evaluative concepts would allow.⁶⁸ Deonna and Teroni clearly expound this second point. In the same way that

we can visually discriminate thousands of shades of colour for which we simply lack the corresponding concepts [...] we may surmise that the sensitivity to evaluative properties that [emotions] authorize is more fine-grained than the discriminations that evaluative judgments provide for. The idea would be, for instance, that the intensity of one's fear co-varies with the degree of danger one faces, something our comparatively coarse evaluative judgments may prove unable to capture.⁶⁹

To reformulate, the suggestion is that emotional experience is similar to perceptual experience in the following way: it can involve discriminations of evaluative properties that are more fine-grained than our existing evaluative concepts would allow, just as visual perception allows us to discriminate between more shades of blue than we currently have concepts for.

This claim seems intuitively true, insofar as it accurately captures our experience of emotion. But there is a further reason to accept it. The fact that emotional experiences can involve such fine-grained evaluative discriminations helps explain why emotions play an essential role in the process of acquiring and mastering evaluative concepts.⁷⁰ Emotions could not play this widely-acknowledged developmental role unless the evaluative nuances offered by emotional experiences could be more subtle than those allowed by our existing concepts. Again, this seems analogous to the case of visual perception. The fact that we can visually discriminate between many shades of blue prior to having access to corresponding concepts is part of what explains how we first develop and master numerous colour concepts.⁷¹

⁶⁸ This point is made most explicitly by theories that identify emotions with perceptions of evaluative properties. E.g., Döring (2007), Tappolet (2011). For other theories that acknowledge these phenomenological resemblances between emotion and perception, see Deonna and Teroni (2012, 66-67), Elgin (2008, 36-37), and Brun et al (2008, 15). In her general analysis of emotion, Nussbaum (2001, 65) uses language reflecting this analogy.

⁶⁹ Deonna and Teroni (2012, 66-67).

⁷⁰ On the role of emotions in conceptual development, see Deonna and Teroni (2012, 84), Brun et al (2008, 20-21), Tappolet (2011, 126-127), and Jaggar (1989, 166-168).

⁷¹ Peacocke (2001).

What does this mean for anger? The idea is that our experiences of anger can make us perceive or sense injustices that our pre-existing conceptual frameworks did not allow us to grasp. In turn, this emotional sensitivity facilitates the development of more nuanced moral concepts. As with the salience role of emotions, then, emotional sensitivity to fine-grained evaluative differences can not only help us first recognise that injustices are occurring, but can also advance our understanding of those injustices, by enriching our conceptual resources for thinking about them.

To illustrate, consider the example, which Fricker discusses at length, of Carmita Wood. One of Wood's male colleagues continuously subjected her to unwanted sexual advances. Though she felt unable to say why at the time, Wood's deep indignation and bemusement made her sense that she was being treated wrongly. Reflecting on these responses in a consciousness-raising group then enabled her and others to develop the concept of sexual harassment.⁷² Thus, Wood's emotional sensitivity allowed her to recognise that she was being treated wrongly before she could even name the wrong in question. And that in turn helped her acquire the conceptual resources needed to understand this wrong. Armed with the concept of sexual harassment, Wood was better equipped to explain how she had been wronged, and reliably to recognise other wrongdoings of the same type.

In sum, recent developments in the philosophy of emotion yield a theoretical account of two ways in which publicly expressing anger, by inspiring anger in one's interlocutors, helps them recognise and understand injustices when they would otherwise have struggled to do so: first, by rendering certain morally relevant properties salient which they had previously overlooked; second, by yielding perceptions of injustice that are more fine-grained than their existing moral concepts would allow. These two roles can come apart. Someone might already have extremely refined moral concepts, but might fail to apply them correctly because some non-moral properties of their environment are not salient to them. Conversely, someone might be aware of all the relevant non-moral properties but have very crude moral categories, which prevent them from seeing or fully understanding existing injustices.

⁷² Fricker (2007, 149–151).

4.2. Two Historical Illustrations

The foregoing account of publicly-expressed anger's epistemic value has been very abstract. To make it more concrete, let us consider two historical examples of political anger. The first focuses on Frederick Douglass's denunciation of slavery. Before proceeding, note what this example is intended to illustrate. We have already seen, in the previous section, how Douglass used his gestures, tones, and rhetoric to express anger and thereby contagiously rouse his audience to anger. This, recall, was vividly reported in Garrison's testimony, where he described the "extraordinary emotion" Douglass's speech excited in him.

What I want to illustrate now is how the anger Douglass transmitted to his audience altered their perception of slavery in the epistemically beneficial way outlined in the previous section. Ideally, to do so, we would consider Garrison's own testimony regarding how the anger he had acquired from Douglass modified his moral perception of slavery. However, Garrison's testimony is not very precise when it comes to describing exactly how the anger Douglass's speech excited in him changed his felt experience of slavery. He merely indicates that this anger made his perception of slavery's wrongness "far more clear". By contrast, Douglass reports in extremely rich detail how experiencing anger transformed his own perception of slavery. So, to illustrate my theoretical account of anger's epistemic value, I will focus on Douglass's testimony regarding his own anger. By considering how Douglass's anger affected his perception of slavery, we can learn about how being roused to anger by Douglass's angry political speech may similarly have enhanced Garrison's (and other audience members') understanding of slavery's injustice.

With this clarification in mind, consider Douglass's account of when, still a slave, he learned to read:

The more I read, the more I was led to abhor and detest slavery, and my enslavers. [...] I loathed them as the meanest and the most wicked of men [...] Liberty! The inestimable birthright of every man had, for me, converted every object into an asserter of this great right. It was heard in every sound, and beheld in every object. It was ever present, to torment me with a sense of my wretched condition. [...] My feelings were not the result of any marked cruelty

in the treatment I received; they sprung from the consideration of my being a slave at all. It was slavery—not its mere incidents—that I hated. [...] The feeding and clothing me well, could not atone for taking my liberty from me. The smiles of my mistress could not remove the deep sorrow that dwelt in my young bosom.⁷³

Douglass clearly experiences intense anger: he “abhor[s]”, “detest[s]”, “loath[es]”, and “hate[s]” slavery and slaveholders. But he is not just telling us that he is angry. Instead, he is also reporting his felt experience of anger at slavery, how slavery appears from his angry perspective. Accordingly, he stresses how every object’s appearance was “converted” by his anger, and repeatedly employs perceptual language.

What does Douglass’s anger teach him about slavery? First, Douglass’s anger draws his attention to how all things, including nonhuman animals, are free and independent in virtue of an “inestimable birthright”. By *rendering salient* the ubiquity of freedom, then, Douglass’s anger highlights his own degraded status.

In turn, this helps him better understand the nature of slavery’s wrongness. The idea is that it is the *status* of slavery that is unjust, the very condition of depending on an arbitrary master, independently of that master’s oppressive actions. In Douglass’s words: it is “slavery—not merely its incidents”—that is wrong. Indeed, as he reports, by the standards he had previously been accustomed to, Douglass’s masters at this time were relatively kind. Thus, the angry perspective helps him perceive more precisely the ground of slavery’s wrongness, which concerns the relative standing or status of slaves. This is quite a fine-grained understanding of the injustice of slavery and of the value of freedom, which arguably anticipates contemporary conceptualizations of freedom as non-domination,⁷⁴ and which Douglass (as well as his audience) may otherwise have been unable to grasp.

Hence, Douglass’s testimony regarding his anger vividly illustrates how anger can both highlight morally relevant properties of one’s environment (e.g., the comparative independence of other living beings), and enable a more fine-grained understanding of injustice than may have been allowed by one’s prior moral categories (e.g., by making one sense the relational injustice of slavery, domination). We may surmise that being

⁷³ Douglass (1855, 159-161).

⁷⁴ Pettit (1997).

roused to anger by Douglass's angry political speech had a similarly beneficial phenomenological effect on Douglass's audience. This helps us understand more precisely what Garrison may have been referring to when asserting that the anger gleaned from Douglass's speech made his understanding of slavery's injustice "far more clear".

Nussbaum might object that Douglass's anger here is too close to her transitional anger to have much dialectical force against her. But Douglass's anger in fact differs from the standard case of transitional anger. His reference to loathing, abhorrence, and hatred suggests a harsh and fiery form of anger, rather than a calm and restrained one. Moreover, he focuses on the injustice itself, rather than on the remedy. This difference is crucial: because transitional anger characteristically does not dwell on the injustice itself, it is ill-suited to helping us understand the nature and depth of that injustice. Admittedly, though, Douglass's anger here does seem qualified in some respects. In particular, the conative dimension of his anger here does not seem to involve anything like retribution or violence.

Hence, let us turn to Malcolm X for an even more intense expression of rage. In his speech 'The Ballot or the Bullet', Malcolm X denounces black Americans' lack of economic and political opportunities:

We're trapped, trapped, double-trapped, triple-trapped. Any way we go, we find that we're trapped [...] So today our people are disillusioned. They've become disenchanted. And in 1964 you'll see this young black man, this new generation, asking for the ballot or the bullet. That old Uncle Tom action is outdated. The young generation don't want to hear anything about "the odds are against us". What do we care about the odds? [...] When we open our eyes today and look at America, we see America not through the eyes of someone who has enjoyed the fruits of Americanism. We see America through the eyes of someone who has been the victim of Americanism. We don't see any American Dream. We've experienced only the American Nightmare. We haven't benefited from America's democracy. We've only suffered from America's hypocrisy. And the generation that's coming up now

can see. And are not afraid to say it. If you go to jail, so what? If you're black, you were born in jail.⁷⁵

Like Douglass, Malcolm X's speech expresses deep anger. And like Douglass, he is reporting what it is like to be enraged at America's racially discriminatory practices, what one sees when one looks through the eyes of an outraged black American. Unlike Douglass, however, his anger does involve the conative dimension Nussbaum is wary of: it explicitly threatens violence ("the bullet") which might well be retributive, born of a desire simply to retaliate ("what do we care about the odds?"). Let us examine what Malcolm X's expression of deep non-transitional anger tells us about how anger affected his perception of racial injustices. As in the Douglass example, doing so will by extension help us learn something about how being contagiously roused to anger by Malcolm X's fiery political speech may have been epistemically beneficial for his audience.

Malcolm X's rage reveals, in place of the American Dream, a vision of the "American Nightmare". Indeed, just as Douglass's anger draws his attention to the freedom of other beings, so Malcolm X's anger renders salient the pattern of deceit that black Americans have experienced: an economic "Dream" that is not genuinely accessible to them, a political "democracy" that they are effectively excluded from, and a civic freedom that seems no different from "jail".

By putting a spotlight on this pattern of disappointments, Malcolm X's rage foregrounds important properties of the injustice black Americans encounter, properties that otherwise may have been obscured by the ideology of the American Dream. Firstly, seeing the pattern suggests that these exclusions are not accidental or isolated but *systemic*, built into the principal political and economic institutions of American society. Moreover, experiencing racial exclusions as patterned and systemic also conveys a sense of their *inescapability* ("we're trapped"), and, consequently, of hopelessness at the prospect of achieving reform from within the American political system. Thus, via its salience role, Malcolm X's anger yields a fuller understanding of the injustice at hand.

Rendering these properties salient, finally, arguably helps refine prior evaluative categories: it contributes to adjusting dominant conceptions of what kinds of political

⁷⁵ Malcolm X (1964).

agency count as *reasonable*. By highlighting patterns of exclusion, Malcolm X's anger helps make rational sense of the apparently unreasonable attitude which consists in embracing violent action, even when the odds of success are unfavourable. If reforming the American system from within is hopeless, and if one's situation is no better than jail, then violence aimed at putting one's opponents in their place may seem the best option.

In sum, these examples illustrate my theoretical account of how experiencing anger can enhance one's understanding of existing injustices and of the reactions they elicit. When we put this together with the fact that publicly expressing anger can *transmit* anger to one's audience by contagion (as happened with Garrison), this indicates how publicly expressing anger can enhance the moral understanding of one's audience.

4.3. Objections to the Epistemic Productivity of Anger

I have already considered some worries concerning anger's epistemic productivity in Chapter 2, when recommending giving a greater public role to emotionally-charged forms of speech more generally. In particular, I raised the common worry that emotions can be manipulated by skilled orators, who use impassioned appeals to excite misguided emotions. In the case of anger, we are all too familiar with how demagogues sometimes express anger at immigrants or racial and religious minorities, in near-total disregard of facts, and thereby arouse misguided rage in their listeners. In her sociological study *Strangers in their Own Land: Anger and Mourning on the American Right*, for instance, Arlie Russell Hochschild argues that Donald Trump's rise to power was facilitated by his ability to provide an outlet for the anger of white working-class Americans, and by his success at turning this anger against vulnerable scapegoats. His success at doing so, she suggests, was itself enabled by widespread factual misconceptions about these vulnerable groups.⁷⁶ Thus, even if angry discourse can be epistemically beneficial when the expressed anger is fitting or well directed, making space for anger in the public sphere risks opening the door to dangerous expressions of *unfitting* or misdirected anger—especially against a background of political ignorance. And since unfitting anger represents just situations as unjust, or innocent people as

⁷⁶ Hochschild (2016, ch. 15, Appendix C).

responsible for injustice, its public expression seems epistemically misleading, not epistemically useful.

There are several things to note in response. In the first place, unfitting or misdirected anger may sometimes retain *some* epistemic value. Fittingness comes in degrees. While some instances of anger are wholly misdirected—e.g., a white supremacist’s rage at having black neighbours—unfitting anger is often only partly so. Some occurrences of anger, for example, might correctly represent a situation as involving injustice while representing the wrong group as blameworthy for this injustice. If the injustice in question is not known or understood by the society at large, expressing this partly-unfitting anger may contribute to advancing our understanding in the way I have theorised. In fact, Hochschild’s study itself brings out this point. While the anger her subjects express is arguably directed at the wrong agents (e.g., minority groups), Hochschild vividly depicts how listening to their anger rendered salient an underappreciated pattern of hardship (rising joblessness, pollution-related illnesses, environmental decay, cultural vilification, etc.). What is striking is how, notwithstanding their mixed fittingness, these angry narratives effectively bring into view an unjust pattern of broken socioeconomic promises—in a way that, in some respects, echoes Malcolm X’s disillusion with the American Dream. In this way, expressions of partly-misdirected anger can lead us to register and examine previously-overlooked injustices, whose real source we may then identify at a later point. This is simply to say that we should not recoil too quickly at the idea of admitting public expressions of unfitting anger: anger that is partly misdirected may retain some of the epistemic value I have theorised.

What about wholly unfitting anger—e.g., the white supremacist’s rage at having black neighbours—which *is* solely epistemically misleading? We have seen that there may be some strategies internal to democratic public deliberation for mitigating its epistemic harms. In Chapter 3, I examined narratives expressing severely misguided viewpoints, which vilify vulnerable groups, and argued that authoritative state-backed speech that denounces such narratives can go some way towards blocking the harms that they would otherwise enact. There, I focused on how state-sponsored counterspeech can prevent public hate speech from eroding the social standing of its targets. But a similar consideration arguably applies to the epistemic harms involved in deeply misguided expressions of anger: by exposing as ignorant and prejudiced the

factual presuppositions on which wholly unfitting anger is based, authoritative speakers can loosen its influence on the public.

Now, this discursive strategy may sometimes be inapplicable. As Chapter 3 acknowledged, state-sponsored speakers are sometimes the ones expressing the wholly misguided narratives.⁷⁷ Even so, a final and more fundamental response is available. In Chapter 2, in response to the concern about misdirected emotions, we saw that non-emotional forms of discourse are actually liable to a similar problem.⁷⁸ In conditions where reliable information is scarce or complex, one of the most successful forms of sophistry consists in endowing fallacious arguments with the appearance of scientific rationality or expertise. And yet, we standardly do not think that because dispassionate argument can be misused in public discourse, we should forego it altogether.

Admittedly, this last response is an unhappy one for democratic discourse more generally: it concedes that political manipulation, which subverts otherwise-epistemically useful modes of discourse, may be widespread in conditions of public ignorance. In Chapter 6, therefore, I will consider head-on the democratic problem of public ignorance. However, the present response does suggest 1) that the problem at hand is not a special problem afflicting angry speech and 2) that just as we do not reject norms of argumentative reasoning simply because they can be abused in non-ideal conditions, so too we should not abandon the use of anger in political discourse.

But publicly-expressed anger faces further objections. While the previous worry concerned the possibility that anger might be misdirected, another goes further and interrogates the claim that fittingly-directed or apt anger is epistemically desirable. Perhaps it is true that such anger can render certain facts more salient, namely those that are liable to ground injustices. But, according to Pettigrove, empirical research suggests that anger also *obscures* important facts. “When angry, people are more likely to see what they take to be hostile stimuli than they are to notice features of their environment they do not take to be hostile”.⁷⁹ Relatedly, “people who have become angry are generally less responsive to counterevidence”.⁸⁰ So, even when it is directed at real injustices, anger risks yielding a distorted perspective on the world.

⁷⁷ Chapter 3, Section 5.1.

⁷⁸ Chapter 2, Section 4.3.

⁷⁹ Pettigrove (2012, 363).

⁸⁰ *Ibid.*, 365.

However, I am not recommending that *all* public discourse be angry. Rather, I am defending the view that angry discourse should play *an* important role in democratic deliberation. The idea is not that angry discourse should supplant the countervailing perspectives that non-angry speech communicates, but that it should complement them. It enriches countervailing perspectives, bringing into view features of the social context that they typically overlook or struggle to accommodate, but does not replace them altogether. When the dominant ideology masks injustices—as when the ideology of the American Dream concealed structural barriers to black Americans’ social mobility—the angry perspective serves as a crucial epistemic corrective, which casts a spotlight on these injustices. But it is a corrective, not a substitute. Therefore, the fact that anger highlights neglected morally relevant facts while obscuring others is relatively unproblematic: non-angry speech can communicate those other facts.

The reason Pettigrove overlooks the possibility that anger might play such a complementary epistemic role is partly that his focus differs from mine. He is primarily concerned with determining whether anger *as a standing disposition of character* is more virtuous than non-angry dispositions, like meekness. By contrast, I am suggesting that, among other forms of speech, public deliberation should involve expressions of anger that are liable to arouse *episodes* of anger in listeners. While having a standing disposition to anger is incompatible with having a standing disposition not to get angry, experiencing an episode of anger towards x seems compatible with experiencing non-angry attitudes towards x at another time. Episodes of anger need not be, nor typically are, permanent. So, angry speech and non-angry speech might alternate, getting their listeners to see the world in different but complementary lights.

Another objection questions the transmission of anger. My account of anger’s distinctive epistemic contribution depends on the claim that publicly expressing anger can help arouse anger in listeners. I have already cited evidence supporting the phenomenon of emotional contagion. However, one might worry that contagion and the perspective-taking it facilitates have limits, which are particularly salient in the non-ideal conditions I am concerned with. After all, doesn’t resonating with the speaker’s anger require that the audience identify with, or at least trust, the speaker? If it does, my argument may seem inapplicable in non-ideal contexts, marked as they are by deep social divisions.

As it stands, this objection is overstated. Expressing anger is not wholly ineffective in divided circumstances. As we have seen, Douglass's rage successfully induced profound anger in Garrison, a white abolitionist. Furthermore, the transmission of anger is sometimes possible even between individuals with significantly different perspectives. Famously, Republican vice-president Dan Quayle once declared that he had vicariously gained important insights into racial injustice from Malcolm X's fiery autobiography: "I can see the hate that was there; I can see the bigotry; I can see it from his perspective".⁸¹

Nevertheless, once qualified, the objection is problematic. There is presumably *some* point at which social divisions and the attending mutual distrust become too great to allow anger to be contagiously transmitted. If group A deeply distrust group B, so that they doubt the sincerity or competence of members of B, it seems unlikely that they will resonate with their angry speech. For example, when prejudicial gender stereotypes represent women as epistemically untrustworthy, those who accept those stereotypes typically experience women's anger not as a signal of some serious injustice to be investigated, but rather as a symptom of women's supposedly hysterical natures. Instead of resonating with the anger, they experience it as alienating or absurd.⁸²

Now, Chapter 5 will examine more closely how far antecedent trust and goodwill really are required for fruitful interpersonal communication. But for now, I assume—along with advocates of the counterproductivity objection such as Nussbaum—that some degree of pre-existing trust is needed for the effective transmission of anger, and that expressing anger will not increase the amount of trust there is.

What can we say about cases where this required trust is lacking? Firstly, Marilyn Frye notes that historically, distrust has fruitfully been tackled by challenging the demeaning stereotypes that underpin it. In the case of gendered distrust, "the struggles and victories of abolitionists, suffragists, prohibitionists, and other reformers made it relatively safe for women to get angry, publicly, on behalf of great moral causes".⁸³ Thus, taking measures to transform gender roles—e.g., by altering media representations of women—can continue to reduce the distrust which prevents many women's anger from being heard.

⁸¹ Quayle (1992).

⁸² Frye (1983, 90-91); Bourgeois and Hess (2008); Lewis (2000).

⁸³ Frye (1983, 93).

Secondly, even when distrust persists, the possibility of ‘networked’ angry speech weakens its significance. Suppose that C deeply distrusts A, so that A cannot transmit her anger to C. But B trusts A, and C trusts B. In such conditions, A can communicate her angry perspective to B. And B, in turn, can *relay* that angry perspective by expressing it to C. In this way, networked angry speech helps to circumvent the communicative barriers raised by distrust.

Nevertheless, these replies have limits. While they may salvage the epistemic utility of publicly-expressed anger in the long term, they do not fully ease our more immediate worries. Just as the affirmative policies Frye recommends may be slow in effecting change, so too relaying angry speech may be arduous. Contrary to the simple schema outlined above, the chain of networked angry speech can be very long in divided settings. Suppose Garrison tried to relay Douglass’s rage. Though Garrison was less distrusted by slaveholders than Douglass, it seems unlikely that they trusted him enough to resonate with *his* angry speech. And slightly more progressive politicians, who trusted Garrison, might have come to distrust him because of his association with Douglass’s message. Hence, Garrison might have to convey the angry perspective to listeners who are even closer ideologically. Progress could be excruciatingly slow.

What is worse, a final concern is that even among those who trust the speaker enough to resonate with her anger, publicly-expressing anger may have countervailing bad effects. This was the point of Nussbaum’s counterproductivity objection: as we saw in 3.1, Nussbaum warns that expressing anger “increases the other party’s anxiety and self-defensiveness”.⁸⁴ Thus, voicing anger risks impeding subsequent communication and coordination with one’s listeners. Consider two ways in which this could happen.

In the first scenario, P is entirely swayed by the angry perspective. But because the anger was intense, and expressed blame and perhaps openness to violence against P’s social group, P feels too unwelcome or afraid to coordinate with the speaker. Perhaps P is a white man who resonates with Malcolm X’s rage towards racial injustices, but is made anxious by his references to white people as “hypocrites” and “devils”. Call P the ‘anxious ally’.

Alternatively, perhaps Q has imaginatively experienced Malcolm X’s rage, but feels that it does not offer an entirely balanced picture of the situation. However, for the reasons just mentioned, she worries that, were she to mention her lingering doubts

⁸⁴ Nussbaum (2016a, 230).

about the accuracy of this perspective, her interlocutors would shun her rather than deliberate with her. Call Q the ‘sceptical ally’.

Despite its promise, then, the epistemic defence of anger’s productivity remains problematic, as shown in three types of cases. In some cases, the problem is that anger hardly seems to have good epistemic consequences. These are cases where the listener does not trust the speaker enough to resonate with her anger. In the other two types of cases—the anxious ally and the sceptical ally—the problem is that despite its good epistemic consequences, it is unclear whether these consequences outweigh anger’s countervailing bad trust-related consequences. These are cases where the listeners trust the speaker enough to vicariously experience her anger, but (because of that same anger) not enough subsequently to engage in practical coordination or deliberation. In what follows, I will show that we can mitigate these remaining problems by reframing our assessment of anger’s consequences.

5. The Systemic Defence of Publicly-Expressed Anger

Section 4 offered a defence of publicly-expressed anger that focused on the kinds of consequences it can produce. I will now complement this defence by examining which *context of evaluation* is appropriate for assessing anger’s consequences. The context of evaluation, recall, is an account of what alternative things we should be comparing along a given standard. Reconceiving the context of evaluation in light of the democratic ‘systemic turn’ renders more tractable the three worries outlined at the end of Section 4.

5.1. The Systemic Turn Revisited

The systemic turn, which Chapter 2 introduced, designates a popular trend among democratic theorists, which enjoins us to reconceptualise the site of democratic public deliberation. On this view, we should not conceive of democratic debate as occurring within a single unitary arena that strives to include all relevant perspectives. Rather, we should think of public deliberation as occurring within a *system* composed of many

discursive arenas which vary in size, function, and formality, and which are connected more or less directly.⁸⁵

Deliberative sites are extremely diverse. Some arenas, like legislative houses, engage in deliberation whose primary function is to produce coercive decisions, and which tends to be regulated by constraining formal norms (such as the requirement that participants appeal to shared reasons). By contrast, more informal arenas—such as rallies organised by social movements—are characteristically more concerned with helping citizens form considered opinions.⁸⁶

As Chapter 2 noted, the systemic view of democratic public deliberation is attractive on various grounds. First, by emphasising that democratic deliberation need not involve a single assembly which includes all citizens, it helps appreciate how inclusive deliberation might be realisable in mass democracies.⁸⁷ Second, notice that the systemic view of deliberation was implicit in my earlier discussion of networked angry speech. Networked angry speech, remember, occurs when A transmits an angry perspective to B, who then relays it to C, and so on. This process presupposes the existence of a plurality of deliberative arenas (one including A and B, one with B and C, etc.) that are connected in virtue of the fact that their constituencies overlap. Thus, the systemic view has already played a role in my defence of publicly-expressed anger.⁸⁸

Crucially, the systemic view renders more tractable the worries outlined at the end of Section 4, by inviting us to adopt a more adequate context of evaluation. To see this, consider how debates concerned specifically with public anger are often framed. Opponents and advocates of anger often proceed by examining the properties of angry speech and the properties of non-angry speech, and asking: which of the two better instantiates our moral standards? Is King's discourse better at realising justice than Malcolm X's? Nussbaum clearly illustrates this tendency when she asks: "Imagining justice anthropomorphically, should justice get angry at offenders?" Whereas

⁸⁵ Mansbridge et al (2012, 2).

⁸⁶ Ibid., 10.

⁸⁷ Ibid., 1-2.

⁸⁸ My endorsement of the systemic view of deliberation here is consistent with Chapter 2. There, I stated that there are powerful reasons to embrace the systemic view. Second, although I criticised appealing to the systemic view to offset the Shared Reasons Constraint's overexclusiveness, my claim was that "absent further considerations", this appeal would not help. I then argued that emotionally-charged narrative should play a greater role in public deliberation. This is compatible with my present claim: that angry emotions should play an important role in democratic deliberation, but only provided that we *also* embrace the systemic view of deliberation.

Nussbaum answers negatively,⁸⁹ advocates of anger such as Srinivasan respond positively.⁹⁰ But what both sides accept, here, is a particular context of evaluation: they compare an instance of angry speech to an independent instance of non-angry speech, and ask which better realises our ideals of justice.⁹¹

The systemic approach brings into view what is wrong with this framing. From a systemic perspective, what we ultimately care about are the properties of the deliberative system as a whole.⁹² We care, for instance, about whether the system is epistemically valuable—i.e., effective at gathering and processing information—and whether it is practically valuable—i.e., conducive to trust that enables coordination against injustice. Consequently, when evaluating parts of the deliberative system, what fundamentally matters is not whether they exhibit these desirable properties when taken independently, but rather whether they contribute to promoting these properties at the systemic level.

Now, Jane Mansbridge observes that “a single part, which in itself may have low or even negative deliberative quality with respect to one of several deliberative ideals, may nevertheless make an important contribution to an overall deliberative system”.⁹³ Take the partisan debates on display in party conventions. Party conventions are generally not open to all, and their debates exclude some experiences and values. However, when seen as part of a broader system that involves other party conventions, Mansbridge argues, such conventions can make the deliberative system more epistemically inclusive. Having a plurality of partisan conventions enables the articulation and development of significantly different perspectives, which can then confront each other in national contests.⁹⁴

Let us apply these lessons to the context of evaluation for angry speech. The first upshot is that we should be concerned, ultimately, with the justice-promoting properties of deliberative systems. Hence, what we should be comparing to our ideal of justice is not *an instance of angry speech* and *an instance of non-angry speech*. Rather, it is *a system*

⁸⁹ Nussbaum (2016a, 173). See also Pettigrove (2012, 355).

⁹⁰ Srinivasan (2017, 11, 20-21), for instance, disputes Nussbaum’s particular anthropomorphic representation of anger, and suggests instead that anger aptly reflects a commitment to justice. See also Bell (2009, 176-180).

⁹¹ In broader debates concerning deliberation, however, Dryzek (2010, ch. 4) comes close to making the point I am making here, as he recommends that rhetoric be evaluated systemically. Rhetoric here presumably includes angry narratives. Nevertheless, Dryzek says very little about anger, and the example of rhetoric he predominantly discusses is King—a paradigmatic example of someone who eschews non-transitional anger.

⁹² Mansbridge et al (2012, 10-13).

⁹³ Ibid., 3.

⁹⁴ Ibid., 19-22.

in which a significant range of deliberative arenas welcome angry speech, and a system in which all deliberative arenas are governed by norms of non-anger. The correct question is not: should we have Martin Luther King or Malcolm X? It is: should the deliberative system contain some important arenas—rallies, debates, etc.—where speakers like Malcolm X are invited to publicly express their anger? Or should we instead prefer a system where, in all arenas, speakers strive to emulate King’s non-anger?

Second, this systemic context of evaluation creates new possibilities for defending publicly-expressed anger. As we have just seen, although partisan rhetoric excludes certain perspectives when taken in isolation, conventions involving partisan rhetoric can contribute to greater epistemic inclusiveness when considered as part of a broader system. By analogy, even if, on its own, publicly-expressed anger erodes the trust that is necessary for subsequent coordination and deliberation, it does not immediately follow that a deliberative system where some (but not all) discursive spheres welcome angry speech lacks this trust.

So far, I have only shown that once we adopt a systemic context of evaluation, the fact that angry speech taken in isolation can have negative effects does not necessarily entail that a system which gives angry speech a prominent place in some deliberative arenas will produce these negative effects more than one that does not. But this only establishes the abstract possibility that angry speech might not be counterproductive. In the following section, I will flesh out this possibility, by revisiting the three problem cases introduced at the end of Section 4.

Before proceeding, we should pause to note what it might mean, concretely, to have a system where some deliberative arenas welcome anger and others do not. In the case of transitional justice proceedings, Chakravarti and Mihai both recommend a division of labour between criminal trials—where speakers are discouraged from expressing negative emotions like anger—and Truth and Reconciliation Commissions—where speakers are invited to voice their angry grievances.⁹⁵ In the case of formal legislative debates, such a system might distinguish between earlier deliberative stages, where discourse charged with intense anger—such as Carol Moseley Braun’s enraged takedown of a patent on the Confederate flag insignia—is welcome, and later deliberative stages, where the presiding officer upholds more orderly discourse.

⁹⁵ Chakravarti (2014, 44-49); Mihai (2016, 28).

In the informal public sphere, the idea is that there should not simply be rallies, conventions, and debates where speakers (perhaps encouraged by a moderator) refrain from intense expressions of anger. The deliberative system should also include a wide range of rallies, conventions and debates where expressions of deep anger are welcome. Alongside the non-angry televised discussion between King and the segregationist James Kilpatrick, there should also be the Cleveland rally where Malcolm X was invited to deliver ‘The Ballot or the Bullet’, or the Harvard or Oxford Union debates which gave him a platform to voice his intense anger.

It is clear that a critic such as Nussbaum would reject this. As discussed in 3.1, Nussbaum considers anger to be “always normatively problematic, whether in the personal or in the public realm”.⁹⁶ Thus, as we saw, she recommends that TRCs give expression to generosity rather than anger; and, accordingly, she praises the way in which Tutu discouraged resentful narratives during the South African TRC. Similarly, recall that in her discussion of informal political contestation, Nussbaum upholds the non-anger of King and Gandhi as an ideal that agents concerned with promoting justice should strive for here and now.

5.2. The Systemic Evaluation of Anger

According to Mansbridge, different “parts of a system may have relationships of complementarity [...] Two venues, both with deliberative deficiencies, can each make up for the deficiencies of the other”.⁹⁷ Let us consider how this complementary relationship functions in a deliberative system that contains sites where speakers express anger as well as sites where speakers refrain from expressing anger. More specifically, how does it help address the three problem cases outlined in Section 4?

The benefits of the systemic perspective are more obvious in the two cases where, although angry speech successfully communicates certain experiences of injustice, it erodes trust in a way that makes subsequent coordination and subsequent deliberation more difficult. Take the anxious ally, who does not trust speakers in anger-welcoming arenas enough to coordinate with them in the struggle against injustice. Here, the existence of arenas governed by norms of non-anger offers a solution: such arenas give

⁹⁶ Nussbaum (2016a, 8).

⁹⁷ Mansbridge et al (2012, 3).

him the opportunity to engage in practical coordination against injustice. Imagine once more the white man who resonates with the anger Malcolm X expresses at the Cleveland rally but feels unwelcome, because of that anger, to coordinate with participants in this rally. Equipped with the epistemic insights about injustice gleaned from the anger-infused rally, he could then coordinate to oppose those injustices within meetings governed, as King's were, by norms of love and openness.

A similar observation holds with the sceptical ally, who, having vicariously experienced the angry perspective, seeks to engage in further deliberation regarding its accuracy, but is reluctant to deliberate with angry speakers. What she seeks is a venue where she can expect her interlocutors to listen to her thoughts and doubts about racial injustice. Deliberative arenas governed by expectations of non-anger provide just such a venue. Indeed, the meetings, rallies, or church-gatherings organised by King's followers adopted norms of love and non-anger partly to facilitate dialogue between estranged social groups.

In these two cases, the relationship of complementarity takes an *additive* form. Each venue independently performs a necessary function: the anger-welcoming venues provide a distinctive epistemic good effect, by transmitting anger-infused experiences; and the anger-discouraging venues provide the trust and openness that facilitates further deliberation and coordination. Thus, when determining the overall properties of the system that contains sites that welcome anger as well as sites that discourage it, we can add their independent contributions, and thereby appreciate how it is preferable to the system that shuns anger.⁹⁸

Understood additively, however, the complementary relationship between anger-welcoming and anger-discouraging venues is unhelpful with respect to the remaining problem case. This is the case where listeners do not trust the angry speaker enough to resonate emotionally with her angry speech. Here, the problem is that angry speech taken independently fails to produce its distinctive epistemic good effect. Consequently, simply adding the independently-achieved good effects of anger-welcoming and anger-discouraging deliberative venues will not reveal what is advantageous about a system which contains angry speech. To address this last problem case, we must consider the *interactive* dimension of the complementary relationship. This dimension concerns how

⁹⁸ Mansbridge et al (2012, 4) and Mihai (2016, 28-29) emphasise the *additive* complementary relationship. Dryzek (2010, 82-83) *highlights* the interactive form, discussed below.

different parts of a system interact to produce good effects that no part would produce independently.

How might anger-welcoming and anger-discouraging discursive arenas interact fruitfully, and thereby help address the last problem case? One suggestion is that the presence of rallies and debates suffused with angry speech may help speakers in venues governed by norms of non-anger acquire the trust of wary listeners. As 4.3 discussed, part of the reason why expressions of anger can make one seem untrustworthy is the existence of prejudicial stereotypes that associate anger with unreason. In such circumstances, the *contrast* between anger-welcoming and anger-discouraging venues can make speakers in the latter appear more reasonable and trustworthy. And this, in turn, can enable them to convey a perspective that they otherwise could not have communicated.

By way of example, consider the deliberative venues inhabited, respectively, by Lincoln and Douglass. These, James Oakes argues, had contrasting but interdependent roles in the abolitionist struggle: the function of the Congressional debates Lincoln participated in was to effect policy changes; the purpose of venues like *The North Star*—the anti-slavery newspaper published by Douglass and Martin Delany—was to denounce slavery’s injustice without restraint or compromise. Importantly, part of what made Lincoln’s contributions seem reasonable to other members of Congress, Oakes suggests, was the contrast between the legalistic calmness which characterised Congressional debates and the fiery moral discourse promoted in venues like *The North Star*.⁹⁹

In a similar vein, Lewis Baldwin contends that, for many white Americans, part of what made it conceivable to engage with civil rights activists who embraced non-anger was the contrasting rage which animated Malcolm X’s rallies.¹⁰⁰ Malcolm X explicitly acknowledged this: “if the white people realize what the alternative is, perhaps they will be more willing to listen to Dr. King”.¹⁰¹ Without the contrast with the deep rage expressed in black nationalist rallies, many white Americans may have deemed King’s vision of interracial love too extreme to be worth listening to.

Oakes and Baldwin are highlighting a similar kind of interaction between parts of the deliberative system. By altering perceptions of reasonableness, they suggest, the

⁹⁹ Oakes (2007, 90-95).

¹⁰⁰ Baldwin (1986, 408).

¹⁰¹ Cited in Baldwin (1986, 398).

presence of anger-welcoming sites helped speakers in anger-free sites articulate their non-angry viewpoints and receive uptake in a way that would otherwise have been impossible.¹⁰²

Now, one concern is that, in this interactive relationship, the viewpoints whose successful communication is facilitated by the presence of anger-welcoming deliberative arenas are not the angry perspectives themselves. So it seems that the distinctive epistemic value of anger is lost. An initial reply to this concern is that the distinctive information conveyed by the angry perspective need not be entirely lost in the promoted non-angry perspectives. A message infused with grief might serve to highlight the loss that results from injustice. And a message of unconditional love, like King's, might mention the existence of injustice, insofar as what makes it loving is its willingness to move beyond a history of oppression and to embrace its perpetrators.

Nevertheless, it is undeniable that some of the distinctive information carried by the angry perspective is lost. Because they are different emotions, grief and love highlight different evaluative properties than anger. Grief may highlight the loss resulting from injustices, but it does not highlight the injustices themselves. And while a loving message may mention injustice, it does not put a spotlight on those injustices or on their perpetrators' responsibility. For instance, the point of King's passing reference to the "sweltering oppression of Mississippi", in his Dream speech, is less to communicate a sense of *what it is like to be oppressed*, and more to introduce a vision of *what it is like to move beyond the divisions which produce oppression*.

Hence, examining the interactive relationship between anger-welcoming and anger-discouraging venues in the deliberative system only partly defuses the challenge posed by the last problem case. In this case, the listener will not resonate with the angry perspective and will not experience its distinctive phenomenology. However, this qualification should not obscure our main result: that *even in this case*, welcoming anger in some deliberative sites still has the potential to produce an epistemic benefit, albeit indirectly. To reiterate: the presence of anger-welcoming sites may facilitate the communication, in other sites, of valuable non-angry perspectives, such as King's loving vision.

¹⁰² For discussion of how Malcolm X and King, as well as Douglass and Lincoln, facilitated this by strategically distancing themselves from each other, see Baldwin (1986, 402-403) and Oakes (2007, xvii-xviii).

By way of concluding this systemic evaluation of angry discourse, an important qualification is needed. When exactly the productive interdependencies I have outlined obtain is, at bottom, an empirical matter which cannot be settled definitively without input from social science. Accordingly, while the examples of the abolitionist struggle and of the Civil Rights Movement arguably illustrate such interdependencies, my point is not that they will definitely always obtain. Rather, as I explained in 3.3, I am making the weaker but nonetheless important point that switching to a systemic context of evaluation renders the productivity worries raised in 4.3 distinctly more tractable.

Recall how it does so. By considering anger and non-anger in isolation from the deliberative system, Nussbaum forces a choice upon us: either we can have angry speech's distinctive epistemic benefits; or we can have non-angry speech's practical benefits for communication and coordination. But we cannot have both. The conceptual shift recommended by the systemic approach reveals that this is a false dilemma. Indeed, the systemic approach brings into view the possibility of complementary relationships, additive or interactive, between spheres that welcome anger and spheres that discourage it. In fact, we have seen that some of the benefits Nussbaum credits solely to non-angry discourse (e.g., communicating a loving perspective) may even depend on such complementary relations.

Thus, invoking the systemic approach allows us to say this: even if Nussbaum were right that angry speech taken alone has a corrosive effect on trust which impedes important types of practical coordination and communication—an assumption I revisit in Chapter 5—it would not necessarily follow that giving anger a significant place *in the deliberative system* requires foregoing non-anger's practical benefits. Indeed, the division of labour between different arenas means that not every deliberative arena has to fulfil every justice-promoting function. To justify shunning public expressions of anger as categorically as Nussbaum does, the counterproductivity critic would have to show that anger-welcoming and anger-free sites cannot complement one another in this way. This is something they have not shown, and which—on the evidence supplied by past struggles for justice—seems highly contentious.

6. Conclusion

I have offered a two-pronged defence of the claim that there is an important place for public expressions of anger inside the deliberative system. Those who deem expressions of anger counterproductive typically overlook anger's distinctive epistemic contribution, and adopt the wrong context for assessing its consequences.

Publicly expressing anger matters partly because, in virtue of anger's particular phenomenology, it can produce a distinctive kind of epistemic good. As philosophical research on emotion indicates, experiencing anger helps render morally significant facts salient and contributes to enriching our moral concepts. Therefore, insofar as it induces listeners to imaginatively experience anger, expressing anger enables them to register overlooked injustices, and to develop a finer understanding of those injustices. In this, we have a systematisation of the compelling suggestion with which we started: that expressing anger is there to teach the audience something, by casting a spotlight on "what [is] really happening". This function matters greatly in non-ideal settings, where members of social groups are often radically ignorant of the injustices endured by other groups.

Taken alone, this first defence encounters several worries. But the force of these worries softens once we adopt a more adequate context for evaluating the consequences of public anger. In line with deliberative theory's systemic turn, we should not ask whether isolated expressions of angry speech have better consequences than isolated expressions of non-angry speech. Instead, the relevant question is whether a deliberative system that gives a key role to angry speech (among other kinds of speech) is more productive than a system that does not. By bringing into view various complementary relationships which can obtain between anger-welcoming and anger-free venues, the systemic perspective renders the above worries more tractable. This second part of my defence captures another intriguing claim with which we began. We now have an answer to why it matters that Ellison's invisible narrator be "dedicated and set aside" for the purposes of expressing his anger: anger is supposed to play specific functions within a broader system, which complement and depend upon other forms of speech.

This philosophical defence has implications for future empirical research on publicly-expressed anger's (counter)productivity. First, given the foregoing account of

angry speech's epistemic value, social scientists should further investigate under what precise conditions publicly expressing anger tends to rouse listeners to anger.¹⁰³ Second, since the systemic perspective reminds us that we cannot conclude much from evidence regarding the effects angry speech has in isolation, they should study more fully the additive and interactive relations of complementarity that may obtain between angry and non-angry speech. In the meantime, I have offered some preliminary evidence for thinking that this avenue of research is promising. In particular, by analysing how it illuminates historical pairings such as Malcolm X and King, and Douglass and Lincoln, I hope to have given a sense of how, when skilfully channelled within the deliberative system, rage offers powerful tools for combatting injustice.

¹⁰³ For preliminary research, see Bourgeois and Hess (2008).

PART 3
ATTITUDINAL
PRESUPPOSITIONS OF
PUBLIC DELIBERATION

CHAPTER 5

DELIBERATION WITHOUT GOODWILL

1. Introduction

1.1. The Attitudinal Presuppositions of Democratic Deliberation

What should democratic public deliberation be like for it to constitute an effective means of realising justice in conditions marked by widespread social divisions and injustice? The past three chapters have defended a set of discursive norms that are geared towards advancing justice in such non-ideal conditions. To prevent political power from being exercised in an arbitrary or dominating way, participants in formal deliberative spheres should appeal to shared considerations.¹ But to offset the exclusionary tendencies of this first norm, public deliberation should also leave significant room for non-argumentative and emotionally-charged forms of speech.² This last suggestion might seem overly inclusive, by opening the door for angry narratives and eventually hateful narratives. However, we should not wholly reject this implication. Angry public narratives perform a crucial epistemic function in divided democracies, by highlighting persistent injustices which may otherwise have been overlooked or misunderstood.³ As for public hate speech, there are strong (though defeasible) grounds for thinking that we should authoritatively denounce it, rather than legally suppress it.⁴

The relevance and effectiveness of these discursive norms presupposes the existence of some non-ideal conditions, including: that many citizens experience systemic injustices; that citizens are often ignorant of each other's experiences; and that some citizens embrace hateful views regarding others. In offering a defence of discursive norms whose desirability depends on such circumstances, my aim has been to supply a normative framework that is action-guiding in our world.

But at this point, one might question just how close my normative framework is to actual, non-ideal conditions. What symbolic and material background conditions already need to obtain for the deliberative norms I have advanced to be realisable and

¹ Chapter 2, Section 2.

² Chapter 2, Section 4.

³ Chapter 4.

⁴ Chapter 3.

conducive to advancing justice? In particular, how far does their adequacy presuppose attitudinal dispositions that are lacking in real-world conditions? In the final two chapters, I consider two attitudinal dispositions. The present chapter examines the disposition to demonstrate trust and goodwill towards one's interlocutors. Chapter 6 will then investigate the disposition to seek and accept reliable information.

1.2. Outline

According to an influential objection, justice-promoting deliberation requires significant goodwill between deliberators. This goodwill, it seems, is lacking in contemporary democracies, marked as they are by deep distrust. Consequently, unless inclusive deliberation plays a role in generating the missing trust and goodwill—a suggestion that this critique typically also rejects—democratic deliberation seems largely unhelpful in actual conditions.⁵

This objection is forceful, and certainly should inform deliberative theories like my own. Nevertheless, the present chapter will develop theoretical resources to help mitigate it as far as possible, and thereby stave off its pessimistic conclusion. After presenting the problem of goodwill (Section 2), I mount a two-faceted attack on its premises. First, I qualify the assumption that deliberation demands high levels of goodwill, by drawing together observations made in Chapters 2 and 4. Specifically, I argue that the division of labour between different arenas in the deliberative system alleviates the motivational demands made on the vast majority of citizens, and helps circumvent these demands when they remain too high (Section 3).

Even with this qualification, however, the objection remains problematic. The second part of my argument therefore pushes back against another assumption: that trust and the attending goodwill cannot arise via non-ideal deliberation (Section 4). Specifically, I investigate how aspects of public deliberation that are especially widespread in non-ideal conditions—namely public hypocrisy (4.1), anger (4.2), and (most controversially) the occurrence of public hate speech (4.3)—have properties that can be made to contribute to generating trust. In saying this, I am *not* claiming that non-deliberative trust-building measures are not needed. Indeed, the trust-building strategies I theorise often involve a mutually-supportive combination of deliberative and non-

⁵ E.g., Johnson (1998), Shapiro (2003, ch. 1), Parvin (2015).

deliberative tools. My point is the more modest one that inclusive discussion remains useful even where antecedent trust and goodwill are largely lacking, because it offers important resources for generating trust and goodwill.

Two preliminary observations are needed. The first concerns empirical evidence. The latter part of my argument considers the potential trust-building effects of deliberation. One might think that this is simply an empirical question, which should interest social scientists rather than philosophers. However, when social scientists gather data, they rely upon theoretical hypotheses that guide their research. As I stressed in earlier chapters, the conceptual insights developed by philosophers can contribute valuably to articulating and enriching such theoretical hypotheses. In this chapter, then, I articulate hypotheses informed by theoretical and philosophical work on topics such as trust, anger, and hypocrisy. At the same time, I will strive to indicate what empirical assumptions these hypotheses depend upon, how far these are consistent with existing evidence, and where—in light of these hypotheses—further empirical research may be needed.

The second point is that the features of democratic discourse which I will suggest have potential for generating trust and the attending goodwill are consonant with the discursive norms that, on independent grounds, I defended as means of promoting justice. Therefore, this chapter's defensive exploration will lend further support to the picture of deliberation that I have been advancing.

2. The Problem of Goodwill

The problem of goodwill is by no means new, and has prominently been articulated by James Johnson and Ian Shapiro.⁶ But it owes its most sustained and general formulation to Phil Parvin, whose argument can be reconstructed as follows:⁷

- (1) Democratic public deliberation that is effective at advancing justice presupposes a high level of goodwill between participants, which itself cannot obtain unless they trust each other to a significant degree.
- (2) Contemporary democracies involve low levels of mutual trust.

⁶ Johnson (1998, 173-174); Shapiro (2003, 22-30).

⁷ Parvin (2015).

- (3) Democratic public deliberation is ineffective at building trust and the attending goodwill.
- (4) So, democratic discursive processes will not be effective at advancing justice until greater trust arises, and will not play an important role in building that trust.⁸

To illustrate, consider a concrete way in which deliberation requires goodwill. As Parvin observes, many accounts of deliberative democracy and of its contribution to justice, including Rawls's or Joshua Cohen's, require that deliberators be civil towards one another. Civility, here, means respecting one's interlocutors by appealing only to considerations that are shared or public. In Chapter 2, I defended a similar norm, which I referred to as the Shared Reasons Constraint. According to Parvin's interpretation, there is evidence that Rawls and Cohen want the norm of civility to regulate the deliberation of ordinary citizens and public officials alike.⁹

This demand for civility, Parvin suggests, is problematic in non-ideal conditions. It seems unlikely that citizens will be civil unless they trust one another to reciprocate: plausibly, I will not commit to respectfully constraining the claims I make *unless* I feel that I can rely on my interlocutors to also fulfil a commitment to discursive civility.¹⁰ But actual levels of trust seem too low to sustain this civility. For Parvin, "increased inequality, social fragmentation, and a weakening of traditional associational life in contemporary states has resulted in a weakening of common bonds of citizenship".¹¹ In the US, for instance, racial tensions between black and white Americans are high,¹² and the affective dislike between Republicans and Democrats has increased dramatically

⁸ I focus on whether participants in deliberation have enough mutual goodwill to conform to certain discursive norms. But Parvin (2015, 408-409) is also sceptical that they will participate at all. Put differently, he is concerned with the rate as well as the quality of deliberation. I only examine the quality issue, as there is evidence that rates of engagement in political discussion are actually quite high. In their comprehensive study, Jacobs et al (2009, 4) find that "citizens engage in more extensive and meaningful public talking than previously suspected". See also Neblo et al (2010). According to Chambers (2012, 63n3), previous studies (e.g., Mutz 2006) reported different findings because they adopted an inadequate operationalisation of deliberative participation. Nevertheless, these recent studies find that the *quality* of deliberation is low. E.g., Jacobs et al (2009, 4), Searing et al (2007, 612). Thus, the pressing question seems to be less whether mass participation is possible, and more whether mass participation *of the right kind* is possible.

⁹ Parvin (2015, 409-412).

¹⁰ Here, I am assuming Hawley's (2014, 9-12) account of trust, where A trusts B to x when A believes B to be committed to x -ing, and relies on B to x . There are competing accounts of trust: for instance, others suggest that it involves relying on others to x *out of goodwill* (e.g., Jones (1999, 68-69)), or out of a *lack of ill-will* (e.g., Lenard (2012, 23)). My arguments here will not hinge on these differences.

¹¹ Parvin (2015, 414-415). See also Putnam (1995).

¹² Massey and Denton (1993, 91-96).

since the 1960s.¹³ What is worse, since (according to the present objection) non-civil democratic public debate is incapable of producing trust and the attending civility, we will need to generate the missing trust via policies that are external to deliberation.¹⁴ In this vein, Elizabeth Anderson defends expansive reforms aimed at rebuilding trust between black and white Americans, which include creating racially-integrated neighbourhoods, restructuring school practices, and adopting aggressive affirmative action policies.¹⁵

Having illustrated the objection, it is worth emphasising its significance. The primary conclusion is that without trust and goodwill, democratic discursive processes are largely ineffective both at advancing justice *and* at generating the antecedent trust that would enable them to advance justice. A corollary of this is that until we build that trust, our instruments for generating trust and advancing justice will be non-deliberative.¹⁶ And since we are allegedly far from having adequate levels of trust and goodwill,¹⁷ this means that inclusive public deliberation will not be a useful instrument in the near future. Given that I am concerned with how inclusive public deliberation can advance justice *in non-ideal conditions like our own*, this last upshot is devastating.

How can we avoid these upshots? It is undeniable that significant non-deliberative trust-building strategies are needed. But this implication alone is not overly problematic for defenders of democratic public deliberation. Archon Fung, among other deliberative theorists, has long acknowledged that in non-ideal conditions, deliberation must be supplemented by non-deliberative practices.¹⁸ Instead, what we must show is that inclusive deliberative processes *also* have a part to play in non-ideal circumstances. To do so, I will challenge premises (1) and (3) in turn.

3. Against Premise (1): What Does Democratic Public Deliberation Really Presuppose?

The systemic understanding of public deliberation helps appreciate that justice-promoting deliberation is less demanding in terms of trust and goodwill than it initially

¹³ Iyengar et al (2012, 412-421).

¹⁴ Parvin (2015, 418-420).

¹⁵ Anderson (2010).

¹⁶ Parvin (2015, 420).

¹⁷ Ibid., 414-415.

¹⁸ Fung (2005). See also Mansbridge et al (2010).

appears. By embracing this more sophisticated understanding of deliberation, we can therefore substantially weaken premise (1). Since the argument here draws together observations made in Chapter 2—when evaluating the Shared Reasons Constraint—and Chapter 4—when examining trust-based obstacles to the contagious transmission of anger—my discussion will be relatively brief.

The systemic approach to deliberation, recall, claims that deliberation occurs in a system composed of many arenas, which differ in size, function, and formality, and which are connected more or less closely. Why does this matter? Firstly, as I observed in Chapter 2, different discursive norms may be appropriate in different deliberative arenas. Participants in formal deliberative arenas, whose primary function is to produce coercive decisions, should appeal to reasons that are suitably public or shared (at least in the final stages of deliberation). But this requirement needn't apply to informal arenas (like newspaper opinion sections) whose main purpose is to help participants form considered opinions.¹⁹ Thus, different spheres needn't be equally demanding in terms of goodwill and trust. In particular, the most demanding norms—like the Shared Reasons Constraint, or as Rawls calls it, 'civility'—may well be required only of public officials taking part in formal deliberations. Hence, even if Parvin were right that some deliberative theorists hold ordinary citizens to standards of civility,²⁰ this criticism is ad personam: deliberative theorists can and should recognise the appeal of civility without demanding it of all citizens.

Nevertheless, further problems arise. Parvin might reply that even in arenas that do not involve strong constraints—such as civility constraints requiring participants to appeal to shared reasons—fruitful public discussion *still* requires excessive goodwill. If so, the problem of goodwill also applies to accounts of public deliberation or to arenas of deliberation that involve less constrained forms of communication.²¹ Why might this be?

Inclusive public deliberation, I argued previously, matters partly as a means of holding one's political opponents accountable, by making them responsive to one's concerns.²² But for inclusive public debate to achieve this, one's interlocutors must be willing to listen to one's complaints. Furthermore, if one's interlocutors listen but are

¹⁹ Chapter 2, Section 3.3.

²⁰ See, however, Chapter 2, Section 3.3, for evidence contradicting Parvin's interpretation of Rawls on civility.

²¹ Parvin (2015, 417).

²² Chapter 1, Section 3.1; Chapter 2, Section 2.2.

unwilling to change their minds or try to answer honestly, it might seem unclear in what sense they are more responsive to one's concerns. Call these attitudinal dispositions—to listen, to be open to changing one's mind, and to respond earnestly—the disposition to engage sincerely. Sincere engagement does intuitively seem required for justice-promoting public debate. Accordingly, Iris Marion Young, despite largely rejecting the aforementioned civility constraint, claims that participants must have “a disposition to listen to others, [...] [to] make an effort to understand them by asking questions” and “to be willing to change [their] opinions”.²³

This is problematic from the standpoint of trust. Intuitively, I am unlikely to listen, change my mind, and respond earnestly unless I trust my interlocutors to reciprocate.²⁴ So, in divided societies where citizens distrust each other, even the weaker attitudinal requirement of sincere engagement may be too demanding. Premise (1) still looms.

One might respond to this problem by challenging the value of sincere engagement.²⁵ But for now I want to accept the *prima facie* plausible assumption that sincere engagement is needed for fruitful public deliberation. Instead, my second point is that the systemic approach contributes to weakening premise (1) in another way, which Chapter 4 introduced. The systemic approach underscores how deliberative arenas with overlapping constituencies can interact with each other. Most significantly for our purposes, a deliberative system makes *networked deliberation* possible. Suppose A and C distrust each other too much to engage sincerely, but that A and B trust each other sufficiently, as do B and C. In such circumstances, A can sincerely engage with B in one sphere. In turn, B can sincerely engage with C, carrying over the information and arguments gleaned from her discussion with A.²⁶

Consider how this second point differs from the first. The first was that adopting the systemic approach can *lower* how much goodwill is required in immediate communicative interactions, by waiving the most stringent discursive norms in some spheres. The second is that even when immediate communication remains too demanding in terms of goodwill, we can *circumvent* this problem by having third parties mediate exchanges between distrustful participants. Contra premise (1), then, high

²³ Young (2000, 24-25). See also Mansbridge et al (2010, 66), Schwartzman (2011, 384-387), and Bächtiger et al (2010, 49).

²⁴ Parvin (2015, 417).

²⁵ E.g., Markovits (2006), Runciman (2008).

²⁶ Chapter 4, Section 4.3.

levels of goodwill and trust do not seem necessary for fruitful exchanges, at least not in the long run.

Nevertheless, this last clause ('in the long run') highlights the principal limitation of my second point. In contexts of deep distrust, there might not be a mediator B who is trusted by both A and C. If so, the network of mediators connecting A and C may have to be extremely long, so that the process of exchanging information and arguments becomes excruciatingly slow. In the immediate future, then, distrust and the attending lack of sincere engagement may reduce the effectiveness of public deliberation at promoting justice.

To overcome this problem, we must resort to substantial trust- and goodwill-building mechanisms. But if, as premise (3) asserts, inclusive public deliberation can play no role in this, then its overall usefulness in the near future remains somewhat limited. Thus, although the systemic approach does blunt the force of the problem of goodwill, we should not let things rest there. Instead, we must challenge premise (3). Accordingly, in what follows, I will explore how deliberation supplies tools for rebuilding the trust and attending goodwill needed for sincere engagement.

4. Against Premise (3): Can Deliberation Produce Goodwill?

Why deny, as premise (3) does, that democratic discursive processes can generate trust and the attending goodwill? For some critics, the problem is that very little evidence supports this claim. Shapiro, for example, states that "there is no particular reason to think deliberation will bring people together".²⁷ Parvin goes further, asserting that empirical evidence largely suggests that deliberation is *ineffective* at bringing people together.²⁸ Thus, the common claim that democratic deliberation can foster trust and goodwill²⁹ may appear to be a mere dogma.

In fact, however, the evidence is more mixed than critics suggest. While some of it is indeed negative,³⁰ much of it is not.³¹ For instance, in a recent deliberative poll,

²⁷ Shapiro (2003, 27).

²⁸ Parvin (2015, 418).

²⁹ E.g., Catala (2015, 435-436), Gutmann and Thompson (2004, 20-21).

³⁰ E.g., Conover and Searing (2005), Searing et al (2007, 610-611), Sunstein (2002).

³¹ E.g., Dryzek and List (2003, 10-12), List and Koenig-Archibugi (2010, 103-109), Fishkin and Luskin (2005), Anderson (2010, 123-127). Even Searing et al (2007) who produce some of the evidence Parvin refers to, actually acknowledge that deliberation *can* produce fellow feeling (612).

Luskin et al invited Protestants and Catholics in Northern Ireland to deliberate together on the future of local schools. Though the issue was divisive, they found that deliberating caused members of both groups to regard each other as more trustworthy.³²

Perhaps, then, the problem is less that deliberation cannot build trust and goodwill, and more—as Patti Lenard worries—that the kind of deliberation that can do so is unlikely to arise in non-ideal circumstances.³³ This refined criticism strikes home against many arguments for using deliberation as a trust-building tool. Consider Amandine Catala’s defence of deliberative trust-building. To guarantee that deliberation will generate trust, she recommends “implementing strict rules of deliberation [...]: for example, listening carefully, speaking respectfully, being responsive to others’ contributions”, etc.³⁴ This is unsatisfactory. We are investigating how deliberation can generate the trust that is necessary for sincere engagement. But in explaining how it might do so, Catala *presupposes* that participants have been made to sincerely engage with each other. Thus, she presupposes the very dispositions we are trying to generate.

A related concern might apply to Luskin et al’s deliberative poll in Northern Ireland. Although the participants came from divided groups, deliberative polls involve highly atypical settings: the deliberation occurs in small groups; it often focuses on well-defined, sometimes relatively technical, policy problems; and it is highly structured and facilitated by trained moderators.³⁵ Consequently one might worry that such deliberative experiments do not tell us much about the more typical and larger-scale deliberation that pervades non-ideal societies.

Nevertheless, some of the positive evidence supporting discursive trust-building *does* look at deliberation occurring in non-ideal conditions. Notably, List and Koenig-Archibugi argue that the public debates which culminated in the creation of the International Criminal Court (ICC) played a crucial part in promoting trust and cohesion between participants. Significantly, this case involved wide-ranging deliberation, rather than deliberation occurring in highly controlled and restricted settings: a complex network of formal and informal dialogue occurred between diverse

³² Luskin et al (2014, 132).

³³ Lenard (2005, 375-376).

³⁴ Catala (2015, 436).

³⁵ Luskin et al (2014, 118).

participants, including not just government officials, but also NGOs and activists.³⁶ Moreover, the discussions took place in a transnational setting, where it is typically assumed that, *ceteris paribus*, bonds of trust are weaker than in national settings. Consequently, their case study supplies *prima facie* grounds for thinking that even large-scale, non-ideal discursive processes can have an important trust-building function.³⁷

In citing this countervailing positive evidence, my point is simply to suggest that the empirical question concerning deliberation's efficacy at generating trust is more open and nuanced than critics of deliberative trust-building often claim. Therefore, it is worth exploring further, both empirically and philosophically.

Now, what is lacking in the above positive evidence is a detailed analysis of precisely *what features* of the non-ideal situation were instrumental in generating trust and goodwill. In the ICC case, was trust-building made possible *despite* non-ideal features of the situation—perhaps by the fact that participants had similar socioeconomic backgrounds? Or did non-ideal features also play a contributory role? And if they did, which ones contributed, and how? While some of the evidence regarding deliberative trust-building provides grounds for optimism, this lack of clarity makes it hard to derive much normative guidance regarding non-ideal public deliberation.

To remedy these shortcomings, I wish to investigate what specific features of the public deliberation which characterises non-ideal conditions might play a role in enhancing trust and goodwill. My aim, though, remains modest. Drawing on philosophical or theoretical work concerning topics such as trust, hypocrisy, and anger, I will develop hypotheses concerning what properties of non-ideal public discourse may be useful, and how. Along the way, I will strive to indicate how these theoretical hypotheses square with existing empirical evidence, and where they call for more empirical research aimed at testing and refining these hypotheses.

Thus, my intention is not to say that non-ideal political discourse always helps build trust overall. Clearly, such discourse is capable of damaging as well as improving trust and the attending goodwill. But it *is* to enhance our understanding of which features of non-ideal discourse have potential for trust-building. Armed with such an

³⁶ List and Koenig-Archibugi (2010, 103-109).

³⁷ For more positive evidence pertaining to non-ideal conditions, see Searing et al (2007, 612), Maddison (2015, 1015), and Steiner (2012, ch. 2).

understanding, theorists and social scientists will be in a better position to explore under what conditions public deliberation can be harnessed to play an all-things-considered positive role in building trust and goodwill.

More specifically, I will explore how three features that characterise non-ideal public deliberation—hypocrisy (4.1), anger (4.2), and (counterintuitively) the occurrence of hate speech (4.3)—offer resources for promoting trust and sincere engagement.

4.1. Public Hypocrisy and Trust

Let us start with a pessimistic assumption about existing levels of goodwill and trust. Suppose A and B are publicly discussing policy issues. However, A is not disposed to listen, respond, or change her mind in response to B's claims. In part, this may be because A simply wants to pursue her self-interest, rather than the public interest. But it may also be partly because, even insofar as A believes the right thing to do is to engage with B in a sincere and public-spirited way, she does not trust B to reciprocate.

Even then, provided that there is a widespread expectation that participants should be public-spirited and sincere, public deliberation creates incentives for A to appeal to public or shared values. As Jon Elster famously argues, “there are certain arguments that simply cannot be stated publicly. In a political debate it is pragmatically impossible to argue that a given solution should be chosen just because it is good for oneself”.³⁸ Now, Elster is almost certainly overstating his argument when he claims that brute appeals to self-interest are *impossible*. Such appeals do occasionally occur.³⁹ But the point remains that public scrutiny creates pressure to frame one's arguments in public-spirited terms. Therefore, even if A is solely motivated to pursue her self-interest, the need to gain public support for her favoured policies strongly encourages her to appeal to public values.

Once A does so, it becomes possible to challenge her on these grounds. Suppose a politician publicly defends tax cuts for the wealthy by appealing to the value of freedom. Her opponents can then observe that if she truly cares about freedom, she should support redistributive taxation instead, since severe poverty makes people unfree. This puts the official in a difficult position. If she disregards the public

³⁸ Elster (1997, 12).

³⁹ Gutmann and Thompson (1996, 126-127).

challenge that she is being inconsistent, or answers it in terms of self-interest, she risks appearing incompetent or hypocritical, and therefore losing support. Hence, the pressures of publicity incentivise her to answer the challenge, and to do so in ethical terms.

The idea, then, is that ill-motivated participants in public debate can be induced to *listen* to ethical challenges, *respond* to them, and eventually *change their stance*, by what Elster calls the “civilizing force of hypocrisy”.⁴⁰ Public attention creates pressure for speakers to hypocritically appeal to ideals that they may not believe in. In turn, the desire not to be exposed as hypocrites constrains them into being more responsive interlocutors.

This effect of public hypocrisy is not mere theoretical speculation. Schimmelfennig argues that it is crucial to explaining why the European Union expanded to Central and Eastern Europe in the early 2000s. Although many powerful EU member states felt that enlargement was against their interest, they were rhetorically trapped into accepting the proposal. European integration had long been publicly justified by appeal to the ideology of a pan-European community of liberal-democratic states. Advocates of expansion exposed inconsistencies between their opponents’ past rhetoric of pan-Europeanism and their later opposition to Eastern expansion. To avoid being shamed as hypocritical—a prospect which might have impaired future cooperation with other member states—recalcitrant member states publicly changed their positions.⁴¹

One might object that public hypocrisy nevertheless does not generate the kind of goodwill we wanted. The question has been how to get participants to trust one another, to acknowledge one another as trustworthy, so that they might then be disposed to engage sincerely with each other—i.e., to listen, respond earnestly, and eventually change their minds. But the pressure of publicity on self-interested actors seemingly does not achieve this. After all, A is *hypocritical*, not sincere, in her dialogical

⁴⁰ Elster (1998, 111). See also Chambers (2004, 405) and Dryzek (2000, 46-47).

⁴¹ Schimmelfennig (2001, 48). Naurin (2007) offers contrasting evidence, which he believes warrants scepticism regarding the civilising force of hypocrisy. However, his evidence seems inconclusive. Naurin’s research focuses on business lobbyists. But unlike with public officials, people do not expect lobbyists to act according to public interest: if anything, the opposite is expected. Accordingly, Naurin (2007, 32) himself concedes that “lobbyists presumably have less to lose from being publicly revealed than, for instance, political parties”. Naurin might nonetheless persist as follows: he finds not just that there is no civilising effect of publicity on lobbyists’ behaviour, but that lobbyists are *less* public-spirited in public than in closed debates with public officials. Again, however, it seems possible to explain this observation without abandoning the hypothesis that publicity usually induces a kind of civilising hypocrisy. In private settings, lobbyists try to win politicians over. And since *politicians* face pressures to be public-spirited, lobbyists presumably need to show them that their proposals can be defended in public-spirited terms.

engagement: even if she listens to and answers B's challenges, nothing guarantees that she actually believes the answers she gives. And A can change her publicly-held position without genuinely changing her mind.⁴² Similarly, A is not made *trustworthy* with respect to discursive engagement. While she might reliably engage in dialogue and answer challenges, she does so purely out of self-interest, not out of a genuine commitment to sincere engagement on these issues. Mere reliability, philosophers of trust widely agree, is insufficient for trustworthiness.⁴³ So, public hypocrisy is insufficient to build trust, and relatedly, insufficient to make participants sincerely engage with each other.

An initial reply is that this shortcoming is relatively unproblematic. This is because A remains functionally equivalent to someone committed to sincere engagement. She is forced, by fear of reputational damage, to listen to others, respond, and eventually change her stated position. Furthermore, unless her response is persuasive, she risks losing the support of listeners. So it may seem inconsequential whether or not she believes what she says: the fact that she is forced to act *as if* she were sincerely engaging in a trustworthy manner is enough to make her responsive and accountable to public criticism.⁴⁴ Indeed, it might not matter to Eastern European states seeking EU membership whether member states that eventually assented really had grown to prefer enlargement. What matters is that public hypocrisy forced them to assent.

We can also go further than this initial reply. Besides increasing A's accountability to B, hypocritically engaging in reason-giving might actually, over time, make A committed to sincerely engaging with B. According to Elster, hypocritical public speakers persistently have to defend values they do not believe in, which is a powerful source of psychological discomfort. To reduce this painful cognitive dissonance, individuals may actually come to believe the ideals they defend.⁴⁵ Hence, however hypocritical and self-interested to start with, publicly upholding the ideal of sincere engagement can eventually make A genuinely embrace this ideal.

Some might have doubts about this first possibility, which depends on psychological theories of cognitive dissonance whose explanatory scope remains contested.⁴⁶ But Robert Goodin suggests an alternative and perhaps less controversial possibility. Sometimes, the problem is not that individuals are purely self-interested, but

⁴² Mansbridge et al (2010, 73-74n28).

⁴³ See Hawley (2014), and note 10, this chapter.

⁴⁴ For this argument, see Markovits (2006, 268), Thompson (2008, 504), and Warren (2007, 278).

⁴⁵ Elster (1997, 12).

⁴⁶ For discussion, see Festinger (1957) and Cooper and Fazio (1984).

that they are only weakly committed to certain ideals. If so, Goodin suggests, public hypocrisy helps remind them of the ideals they already embrace, and underscores the implications of those ideals.⁴⁷ This second process requires that participants have *some* measure of pre-existing commitment to the ideal in question. However, this is not exceedingly worrying in our context: there is good evidence that citizens are often at least weakly committed to the ideals of listening to, trying to understand, and answering the claims of other participants.⁴⁸ In such conditions, public hypocrisy can reinvigorate the commitment that very many participants have to sincere engagement, and thereby make them more trustworthy interlocutors.

Now, the present strategy for building goodwill and mutual trust, which proceeds through publicly-induced hypocrisy, does presuppose that some background conditions hold. First, as previously mentioned, there must be a widespread expectation that public speakers should engage in a sincere and public-spirited manner. But this, Elster suggests, is not necessarily unrealistic: it is consistent with many citizens not embracing this ideal, and with some individuals endorsing it but being too weak-willed to comply with it.⁴⁹

Second, the speaker's discourse must be made public via monitoring, so that other participants can impose sanctions when speakers fail to act in a sincere and public-spirited manner.⁵⁰ Even so, in her detailed examination of trust and democracy, Lenard suggests that the sanctions in question needn't be very heavy-handed. They could simply consist of reputational damage, lack of future cooperation, or, for elected officials, the threat of being voted out. As for monitoring mechanisms, Lenard observes that they already "are a regular part of democratic life".⁵¹ For instance: media outlets habitually report speakers' present claims and past statements, thereby enabling listeners to hold them accountable when the two are inconsistent.

Still, one problem with monitoring speakers is public ignorance. Even if reliable media sources monitor public agents, participants may fail to avail themselves of this information. Worse, they might instead defer to unreliable sources. For now, I assume

⁴⁷ Goodin (1992, 137).

⁴⁸ Conover and Searing (2005, 276).

⁴⁹ Elster (1995, 248-249).

⁵⁰ Although I emphasise publicity's value, this is compatible with thinking that secrecy can also play a valuable role. On how to combine public and secret deliberation, see Gutmann and Thompson (1996, 114-117) and Chambers (2004, 405-409).

⁵¹ Lenard (2012, 143-145).

that we can achieve significant levels of reliable monitoring by the public. Chapter 6, however, will consider the more pessimistic scenario, by directly examining contexts where citizens are highly ignorant about politics.

Let us conclude. Public deliberation in non-ideal conditions is rife with hypocrisy. This provides theoretical grounds for thinking that such discourse can generate more favourable dispositions in participants *even if* we start from sober assumptions regarding their initial levels of goodwill and trust. More specifically, public hypocrisy can dispose participants to listen to one another, offer responses, and eventually change their stated positions. Admittedly, this mechanism presupposes some felicitous background conditions, including the existence of mechanisms for monitoring and sanctioning public speakers. As Lenard notes, however, these are often already part of democratic politics. Additionally, if we make further—albeit modest—empirical assumptions (for instance, about cognitive dissonance), there is reason to think that public hypocrisy may over time generate a genuinely trustworthy commitment to sincere engagement. Importantly, however, what follows will not depend on those further empirical assumptions. Instead, I want to examine how, once we grant that participants are at least disposed to listen to each other, other typical features of non-ideal public discourse can contribute to building trust and goodwill.

4.2. Anger and Trust

In divided contexts, an important aspect of public deliberation is the exchange of personal narratives, or testimony, including emotionally-charged narratives. Chapter 2 recommended welcoming such narratives on *epistemic* grounds—to remedy the ignorance of some citizens regarding others. But philosophers concerned with transitional justice frequently emphasise their usefulness for rebuilding trust in divided contexts. For example, citing cases like South Africa’s post-apartheid Truth and Reconciliation Commission—where victims and perpetrators were invited to publicly tell their stories—Cécile Fabre argues that sharing difficult stories regarding what one has endured, often in a way that is suffused with negative emotions, can contribute importantly to transforming damaged social relations.⁵²

⁵² Fabre (2016, 280). See also Chakravarti (2014), Maddison (2015, 1022-1023), Mihai (2016), Steiner (2012, 84). Importantly, the success of TRCs at generating trust is mixed. While some evidence is positive (e.g.,

Crucially, such transitional periods involve extremely low levels of mutual trust and goodwill. Accordingly, drawing on the resources developed within transitional justice debates seems promising when challenging premise (3) of the problem of goodwill: it allows us to investigate how public discourse *in highly non-ideal contexts* might help generate trust and the attending goodwill.

Let us consider what kinds of emotionally-charged personal narratives might be useful, and how. One might think that some kinds of emotionally-charged testimony can contribute to rebuilding trust, while denying that *angry* testimony can do so. For example, as Chapter 4 discussed, Nussbaum explicitly affirms that angry narratives are deleterious for trust.⁵³

But saying this would weaken my reply to the problem of goodwill. First, angry narratives are pervasive in non-ideal conditions, which characteristically involve widespread injustices. Thus, excluding them from discursive trust-building processes represents a significant qualification of my suggestion that typical features of non-ideal public speech can be made to contribute to these processes. Moreover, discouraging angry speech because it allegedly inhibits trust-building procedures jars with my recommendation, in Chapter 4, that we give a substantial place to such speech.

For these reasons, following Sonali Chakravarti's examination in *Sing the Rage: Listening to Anger after Mass Violence*, I focus on *angry* personal narratives.⁵⁴ Chapter 4 defended angry narrative by arguing that it had distinctive epistemic value, while accepting for the sake of argument Nussbaum's claim that its impact on trust and goodwill is at best inert and at worst disastrous. I then argued that even conceding this, an appropriate understanding of public deliberation helps to circumvent anger's alleged bad consequences on trust. Here, by contrast, I will push back against the assumption that the influence on trust of angry narrative is categorically negative.

This is of course not to say that its influence on trust is always positive overall. Establishing when exactly its influence is overall positive remains partly an empirical task. But that task, as I explained in the introduction, depends on theoretical hypotheses regarding what specific positive trust-building effects angry testimony might have. These hypotheses should be informed by the conceptual work philosophers have

Gibson (2004)), some is not (e.g., Shaw (2010)). My aim is to develop hypotheses regarding what might explain the positive side.

⁵³ Nussbaum (2016a, 213, 233).

⁵⁴ Chakravarti (2014, ch. 6). See also Maddison (2015, 1022).

done on anger, trust, and testimony. Accordingly, my aim here is to enrich empirical investigations by developing such hypotheses. So let us ask: assuming that discursive participants listen to one another, how might voicing angry personal narratives build trust between them?

It will help to have a case of angry testimony before us. Chakravarti offers the poignant example of Lendiso Richard Ndumo Galela, an anti-apartheid activist who was tortured while in detention. Galela supplied the following testimony at the South African TRC:

There is no amount of money that can pay for a person's dignity. That is a fact. I am the equivalent of a corpse, I do not have any dignity left. I have been stripped of my dignity and that is what I wanted to say and I would just like to say that even though my physical appearance seems like that of a man I have been stripped of my manhood and a specialist told me that it would give, it would take three months for me to be treated and if nothing happens, I should know that I would be condemned sexually. [...] What I am saying is that I do not see what the Commission can do for me.⁵⁵

Why might listening to this resentful narrative induce the listener to trust Galela? A recurrent suggestion among theorists of transitional justice, which Fabre notably articulates, is that hearing such a narrative can have a humanising influence: it helps the listener see the speaker as a human being, whose humanity is worthy of concern and respect. Anger, recall, represents its target as involving a moral violation or injustice. So, an angry personal narrative reveals that the speaker is grappling with a powerful sense of injustice ("I am the equivalent of a corpse [...] I have been stripped of my dignity"), and discloses the difficult experiences which prompted that sense (how the torture Galela sustained resulted in his impotence). Being confronted with the speaker's psychological agitation and discovering the painful experiences he has endured already helps us understand who he is and where he is coming from. Accordingly, it helps the

⁵⁵ Cited in Chakravarti (2014, 138).

listener perceive the speaker as an individual in their own right—as a human being with particular experiences and concerns.⁵⁶

This humanising influence is important. As Katherine Cramer and Elizabeth Anderson emphasise in their recent examinations of urban/rural divisions and racial divisions in the US, distrust is often driven by the fact that individuals from estranged social groups perceive each other solely as members of a resented outgroup, which erases or obscures their shared humanity.⁵⁷ Insofar as this is the case, angry narratives' humanising function seems promising with respect to trust-building.

But there is another, less examined, way in which listening to anger might be positively associated with trust: listening to angry testimony might enhance the listener's disposition to act trustworthily because, in some cases, the speaker's angry narrative itself constitutes a manifestation of trust. Telling one's story can require disclosing deeply sensitive aspects of one's life, particularly in conflict-ridden circumstances. This seems likely not just when expressing grief or despair, but also anger. Since the appropriate target of anger is a moral violation,⁵⁸ angry testimony may sometimes involve revealing how one has sustained violating or degrading treatment. For Galela, giving testimony means expressing anger at a deeply humiliating injustice ("I have been stripped of my dignity [...] I have been stripped of my manhood"). In such cases—where the injustice that the speaker's anger targets is degrading or humiliating—delivering angry testimony implies baring oneself, and hence making oneself vulnerable to one's listeners. Now, according to philosophical analyses of trust, voluntarily making oneself vulnerable to others *just is* placing trust in them.⁵⁹ Thus, in making himself vulnerable to his listeners, the speaker whose anger targets humiliating wrongs he has sustained is signalling trust in his listeners—in particular, trust that they will not take advantage of that vulnerability by, say, mocking him.

Why does this matter? Pettit argues that when A places trust in B, this generates a new reason for B to act in a trustworthy way and eventually to reciprocate by extending trust. To place trust in another, Pettit suggests, is to show esteem for them. Since individuals typically seek the esteem of others, they desire to maintain this trust. Therefore, placing this trust gives the trustee a new reason, based in their desire for

⁵⁶ Fabre (2016, 264). See also Chakravarti (2014, 153), Maddison (2015, 1021), Halpern and Weinstein (2004, 571-579), and Aiken (2010, 185).

⁵⁷ Cramer (2016, esp. 8-9); Anderson (2010, 44-50).

⁵⁸ Chapter 4, Section 2.

⁵⁹ E.g., Baier (1986, 235), Pettit (1995, 208), Jones (1996, 12), Lenard (2012, 18-20).

esteem, to act trustworthily. This is what Pettit calls the “cunning of trust”.⁶⁰ In making himself vulnerable to his listeners through his anger-infused narrative, then, Galela gives them a new reason to act in a trustworthy way.

One might worry that the “cunning of trust” does not apply to non-ideal circumstances. This property of trust requires that the trustee desires the truster’s esteem. But we often recoil at being esteemed by people we deeply dislike.⁶¹ Consequently, if the speaker and listener come from estranged social groups, one might think that the intended trust-building effect will not occur.

This worry derives much of its force from the ambiguity surrounding what kind of esteem the speaker is signalling to the listener. What makes us recoil from the esteem of our enemies is typically the worry that they extend this esteem for the wrong reasons. I worry about being appreciated by racists because of the thought that they appreciate me *because I am not black*. But, importantly, the kind of esteem being communicated here does not necessarily seem problematic in this way. At least part of what the speaker’s trusting reliance communicates, here, is the expectation *that the listener will not take unfair advantage of their vulnerability*. Intuitively, being respected as the kind of person who does not take unfair advantage of others’ vulnerability is something we have good moral reason to seek, even from those we dislike.⁶²

To illustrate this last point, consider an example. During the 1963 Birmingham protests in Alabama, Commissioner for Public Safety ‘Bull’ Connor commanded his men to target civil rights protesters with fire hoses. The protesters responded by getting to their knees and praying. Thus, they made themselves vulnerable to police forces, and signalled their expectation that the police would not exploit that vulnerability. Struck by this gesture, Connor’s men let the marchers through.⁶³ On one interpretation, the protestors’ act of trust communicated a kind of esteem that Connor’s men had reason to value, even though it came from a group that they otherwise (we can plausibly assume) strongly resented.

So far, I have emphasised how *hearing* testimony charged with anger might dispose the listener to engage in trusting relations. But *delivering* the narrative, Fabre notes, might also help. One might be unwilling to trust others because one feels resentful at being

⁶⁰ Pettit (1995 214).

⁶¹ Ibid., 221.

⁶² This reason may be overridden (perhaps a racist confides in me that she plans to commit a hate crime). But it is a reason nonetheless.

⁶³ Anderson (2010, 100).

ignored or misrecognised. If so, publicly relating what one has endured seems to be part and parcel of what is needed to restore one's trust.⁶⁴ Again, Galela's case is instructive. What keeps Galela from participating in social life is partly the perception that his ordeal has degraded him, and that this degradation is not properly acknowledged ("no amount of money [...] can pay for a person's dignity"). Accordingly, having his terrible experiences and his resulting anger acknowledged seems to play a constitutive role in restoring his willingness to engage in trusting relations. Giving impassioned testimony and receiving uptake, then, can be part of what makes speakers trust their listeners.

A closely related point, which Chakravarti emphasises throughout her analysis of anger in TRCs, is that giving angry testimony may also be *instrumental* to making the speaker trust others. Because angry testimony delivers facts about the world and the speaker's psychological state—and more specifically, about the wrongs they see themselves as having sustained—it yields insights into what citizens need to restore trust.⁶⁵ As we have just seen, Galela's narrative indicates that financial reparations are the wrong kind of response to his anger. Instead, expressive acts aimed at making him feel dignified and respected seem a more fitting remedy.

We have, then, several theoretical hypotheses regarding how testimonies charged with negative emotions such as anger might contribute to rebuilding trust and goodwill. They can help the listener perceive the speaker's humanity; signal trust that incentivises trustworthiness; make the speaker feel recognised as a member of society; and yield information regarding what measures are needed to restore the speaker's trust.

Importantly, although Galela's narrative arguably exemplifies all of these trust-building functions, not all types of angry speech should be expected to perform these functions equally. For instance, the hypothesis that angry speech signals trust in a trust-incentivising way applies especially to cases where the injustice targeted by the anger is degrading or humiliating. To reiterate, this is because these are the cases where it is clearest that revealing the injustice suffered makes one vulnerable, and therefore constitutes a signalling of trust. This restriction needn't apply to the other trust-building functions of angry speech I have outlined. Delivering her testimony might conceivably help the angry speaker feel acknowledged, or yield information regarding how to

⁶⁴ E.g., Fabre (2016, 265), Tirrell (2015, 247).

⁶⁵ Chakravarti (2014, 3, 20-21, 167).

restore her trust, or make listeners appreciate her humanity, irrespective of whether the injustices she is reporting were humiliating.⁶⁶

Thus, far from suggesting that all instances of angry narrative will perform these positive trust-building roles equally, the theoretical hypotheses I have outlined help explain why we can expect different kinds of anger to be effective at different functions. In contrast, existing empirical evidence regarding how angry communications affect interpersonal relations often does not distinguish between different kinds of anger. And when it does, it rarely makes discriminations based on the intentional objects of the anger (e.g., what kind of injustice the anger is directed at, whether it targets the injustice only or also its perpetrators, etc.).⁶⁷ As illustrated above, the philosophical hypotheses I have outlined give us reason to expect that such differences may influence angry speech's impact on trust. Thus, these hypotheses can and should guide the empirical task of ascertaining precisely when angry narratives can have a positive influence on trust, all things considered.

Even with this qualification, objections remain. First, one might worry that the foregoing discussion does not show much about the trust-building effectiveness of angry testimony itself. After all, my running example is taken from the South African TRC, whose discursive processes relied upon heavy institutional support. In particular,

⁶⁶ Another dimension concerns less the *kind of injustice* that the anger is directed at, and more whether the angry speech targets *the perpetrators of the injustice*. Galela's testimony hardly refers to the perpetrators. For instance, he uses the passive tense when saying "I have been stripped of my dignity". Would the trust-building effects of angry testimony decline if Galela's anger targeted the perpetrators ("Look what you've done to me! You have stripped me of my dignity!")? This depends on what particular trust-building function we are examining. With respect to how delivering an angry testimony helps the *speaker* trust the audience, it is unclear why it would make a difference. However, things seem more problematic when considering how angry testimony helps the *listener* trust the speaker. Being targeted by anger might make the listener feel defensive, not trusting. Some evidence seemingly supports this worry. Steinel et al (2008) find that negotiators are more concessive when confronted with anger that targets their behaviour than when facing anger directed at their person. Nevertheless, the study remains inconclusive for our purposes. First, it focuses on what concessions were made, rather than directly measuring perceptions of the speaker's trustworthiness. Secondly, it considers cases where the expressed anger solely targets the behaviour ("This offer makes me really angry"), and where it solely targets the person ("This person makes me really angry"). This leaves it unclear what to make of 'mixed' cases where the speaker's anger explicitly targets the listener *because* of their behaviour ("I'm angry that you made me this offer."). So it remains inconclusive for angry statements like "You have stripped me of my dignity!"

⁶⁷ For studies that do not distinguish different kinds of anger, see, e.g., Tiedens (2001), Sinaceur and Tiedens (2006), Van Beest et al (2008), and Wubben et al (2009). Some studies do discriminate between different kinds of anger, but they typically draw distinctions that cross-cut the ones I suggest are philosophically interesting. Indeed, they mainly distinguish between threatening and non-threatening expressions of anger, or direct and indirect expressions of anger. E.g., Guerrero (2009), Sereno et al (1987). Neither of those distinctions concerns what the anger's intentional object is. One exception is Steinel et al (2008), who distinguish between person-oriented and behaviour-oriented anger. As note 66, this chapter, discusses, however, their research has limitations for our purposes.

they were facilitated by moderators who, *inter alia*, reminded participants of what was expected of them—sincerity, openness, listening, etc.

But even if skilful facilitation by moderators is a necessary ingredient in trust-building, this does not undermine the more fundamental point: that *given* skilful facilitation, negative personal narratives offer important resources for building trust in deeply divided contexts, and (as the example of the TRC suggests) can seemingly do so in large-scale deliberative settings. More generally, this suggests that even if they must be combined with non-deliberative measures, deliberative processes can a) contribute to trust-building within fragmented and unjust societies, b) in virtue of some of their non-ideal properties (here, anger). Limited though this point may be, it still casts doubt on premise (3), the categorical claim that deliberation cannot contribute to trust-building in non-ideal conditions.

Nevertheless, a second worry arises. In exploring anger's trust-building potential, I have focused on fitting instances of anger, where the angry testimony denounces genuine injustices. But what about *unfitting* anger? As discussed in previous chapters, many occurrences of angry speech in non-ideal conditions express unfitting or misdirected anger. For some, publicly revealing their impassioned viewpoint means expressing rage at marginalised groups. So making space for angry storytelling also risks letting in public expressions of xenophobic or racist rage.⁶⁸ Allowing such hate speech, it might seem, is disastrous for trust-building. The worry is not just that such speech cannot perform the aforementioned positive trust-building roles. It is that, in addition, it risks having a highly negative impact on trust.

Chakravarti recognises this worry, and therefore stipulates that angry speakers should have the freedom to tell their story “as long as it does not become hate speech”.⁶⁹ However, this response seems institutionally problematic. We are supposed to make space for personal storytelling, where individuals voice the grievances they harbour deep down. Simultaneously, deeply misguided grievances must be excluded.

⁶⁸ This worry is especially salient with online deliberation, which is rife with disrespectful narratives. According to Levmore, this is because online anonymity enables speakers to express disrespectful views with impunity. But there are ways of counteracting this specific problem. For instance, as Moore (2017, 20–21) reports, requiring online participants to supply their real names—or simply to maintain the same pseudonyms over time—drastically increases the respectfulness of their contributions. This suggests that online discourse can be made less *distinctively* problematic with respect to trust-building (as compared with offline discourse). For critical discussion of the claim that online disrespectful speech is distinctively problematic, see Brown (2017).

⁶⁹ Chakravarti (2014, 156).

But, in context, part of what is disputed *just is* which grievances are deeply misguided. Therefore, it seems difficult unambiguously to stipulate which views are ‘out’ before hearing people’s testimony. And given this uncertainty, excluding hateful views risks leading to self-censorship that defeats the aim of having people say what is really troubling them.

Perhaps there *is* an institutional way of avoiding this problem. But I do not want to rely excessively on this possibility. Moreover, Chapter 3 introduced defeasible reasons against forcibly or legally excluding public hate speech. Hence, considerations of consistency militate against taking this route. Finally, even if we could suppress hate speech without inhibiting the expression of fitting angry testimonies, this would constitute a significant concession to the problem of goodwill: it concedes that we need heavy-handed sanitisation of public deliberation before it can be used for trust-building. Instead, then, the final section will introduce reasons for doubting the view that failing to forcibly exclude hateful public narratives would necessarily undermine the process of building trust and goodwill.

4.3. Hate Speech and Trust

Briefly put, the concern is that public hate speech seems likely to alienate listeners, particularly those it targets. Hate speech denies its targets’ basic social standing. Therefore, if A utters hate speech which targets B, this will presumably drastically diminish the trust and goodwill B has towards A. Now, failing to exclude hate speech increases the opportunities for publicly expressing such speech. From these premises, one might conclude, failing to exclude hate speech is bound to have a destructive effect on trust and goodwill.

But this argument is too quick. Despite accepting its premises, Marcus Schulzke reaches the opposite conclusion: legally permitting hate speech, he contends, is in important respects instrumental to rebuilding social trust. The crux of his argument is that, because allowing hate speech increases opportunities for publicly expressing hateful views, doing so helps identify who does and does not hold hateful views.⁷⁰ This identifiability, Schulzke suggests, matters for two trust-related reasons.

⁷⁰ Schulzke (2016, 233).

Firstly, it makes hate speakers and their beliefs easier targets of counterspeech. Once hateful views are in the open, we can publicly repudiate them, and expose their baselessness.⁷¹ Although this might conceivably get hate speakers to abandon their abhorrent views, Schulzke concedes that this is unlikely. More significant is the potential effect of counterspeech on third parties. “Bystanders, especially those who might be susceptible to becoming prejudiced, can profit from seeing public debate [...] that demonstrates how little evidence there is to support hateful belief systems”.⁷² Persuading third parties that bigoted views are misguided through reasoned debate is not altogether uncommon. In 1965, for instance, civil rights activist James Baldwin publicly debated conservative thinker William Buckley (a man who famously referred to whites as “the advanced race”). That day, Baldwin’s eloquent logic easily won his audience over. Examples like these suggest that allowing public hate speech creates opportunities for counterspeakers to change the minds of those who might have been tempted by degrading views. This, in turn, seems good for trust: repudiating hateful views makes one more capable of forming trusting relations with the targets of those views.

Such examples notwithstanding, Schulzke’s first trust-based argument for allowing hate speech needs refinement. Why not think that the greater public presence of hateful views will draw *more* people to those views? Schulzke assumes that because hateful views are unsupported by good reasons, counterspeech will generally show them to be unpersuasive. But how persuasive a view appears is not wholly determined by the logic supporting it. The perceived authority of the person voicing it also matters. What is needed, then, is not just for private citizens to respond to hate speakers. Instead, as I argued in Chapter 3, authoritative public officials, and ideally state-backed officials, must actively denounce public hate speech.

Even with this refinement, there are limits to what counterspeech can achieve by way of persuasion.⁷³ Some hateful citizens will be unwilling to change their minds. Worse, feeling publicly challenged might further entrench their commitment to hateful views. So even if Schulzke is right that the identifiability of hateful doctrines creates new opportunities for steering some citizens away from hateful views—and, in doing

⁷¹ Ibid., 236. See also Brettschneider (2012).

⁷² Schulzke (2016, 236).

⁷³ This is compatible with Chapter 3’s argument: there, I focused not on how authoritative counterspeech *persuades*, but rather on how it *blocks* hate speech’s assault on its targets’ dignity.

so, helps to improve social trust—it nonetheless seems to have some countervailing bad effects on trust: in particular, the speech of hateful citizens who do not abandon their views will continue to alienate targets of hate speech.

This is where the second and more important trust-related function of making hate speakers identifiable kicks in. To appreciate this function, consider that even if being on the receiving end of hate speech makes trusting the hate speaker impossible, it does not necessarily follow that things would be better with respect to trust if hateful citizens were forced *not* to speak their minds. There are two possibilities. Either hate speech bans make the existence of hate speakers wholly invisible, or they do not.

If hate speech bans utterly conceal the existence of hateful citizens, they put targets of hateful views at risk of engaging in *unwarranted* relations of trust—that is, of trusting citizens who are untrustworthy. Suppose that B does not know that A is a hateful citizen, and, out of ignorance, relies on A. The trust B places in A seems unwarranted: A's hateful belief that B does not belong in the community of equals to which A is accountable seems inherently in tension with a commitment to behaving trustworthily towards B. Now, as Russell Hardin explains, engaging in such an unwarranted relation of trust is undesirable: because unwarranted trust is by definition likely to be betrayed, it is both unstable and dangerous for the truster.⁷⁴

The second (and more empirically plausible) possibility is that hate speech bans do not make hateful citizens wholly invisible. Even if hateful citizens do not publicly express hateful views, it is still known *that some people hold hateful views*. But what becomes less well known, Schulzke emphasises, is *who* exactly holds these views. This lack of identifiability, he argues, is again problematic. Insofar as they do not know who holds

⁷⁴ Hardin (2002, 29-32). Two points regarding consistency are worth flagging here. Firstly, this last argument is compatible with my claim, in 4.2, that placing trust produces a new incentive to be trustworthy. The difference is that, here, B places trust *out of ignorance*: the description of A which motivates B to extend trust is one that A does not actually satisfy. Consequently, A cannot justifiably think that B is signalling esteem for A. Second, one might worry that my more general argument in 4.3 is inconsistent with my discussion of the civilising force of hypocrisy in 4.1. Section 4.1 argued that informal norms can induce hypocritical behaviour in ill-motivated participants, which in turn can 'civilise' them. But if hypocrisy has such a civilising effect, why not *also* impose formal legal restrictions on hate speech, with the aim of further incentivising hypocrisy in hate speakers? There are several reasons to resist this recommendation. I have already introduced the first: *prima facie*, legally restricting hate speech risks interfering with making space for angry testimony aimed at advancing trust. Additionally, in 4.1, the civilising force of hypocrisy was used more modestly: a) it applied to individuals who were typically less ill-motivated. The paradigm cases were egotists and people who have decent values but are weak-willed. Hate speakers, by contrast, have morally objectionable views. And b) the aim of the civilising force of hypocrisy was less ambitious. It was to get ill-motivated citizens to engage sincerely with others. This seems less demanding than getting individuals to revise vilifying prejudices, which often resist evidence. While vilifying stereotypes might sometimes be what prevents individuals from engaging sincerely with others, there is no reason to think that this is generally so. Because of these disanalogies, the civilising force of hypocrisy may be less applicable here than in the cases discussed in 4.1.

hateful views, targets of hateful views must either run the risk of trusting untrustworthy hateful citizens, or refrain from trusting anyone from groups that are known to contain some hateful citizens.⁷⁵

To illustrate, consider Zimasile Bota's testimony to the South African TRC. In her testimony, Bota asserted:

One of my requests is that the perpetrators, the Government must bring them forward. We cannot work together with the police. We know that the police are protecting us, but we cannot work together with the police who have blood on their hands. [...] My request is that the police must come forward.⁷⁶

Bota is claiming that because many police officers were complicit in racist apartheid practices, they cannot be trusted. Until those who were complicit are identified, working with the police seems impossible, as it is unclear which members are trustworthy. Restoring trust, she suggests, requires making it more visible who is and who is not trustworthy.

This, Schulzke emphasises, is precisely what legally permitting public hate speech, while advocating (as I have done) robust counterspeech against it, would in principle achieve.⁷⁷ Allowing A to voice his hateful views towards B, while having non-hateful third parties (C) denounce these views, helps B see that while she cannot trust A, she can trust C. This seems intuitively preferable to the two previous scenarios: 1) the situation where B does not know that anyone has hateful views, and therefore misguidedly trusts A; and 2) the situation where B only knows that one of A and C has degrading views, and is therefore unsure which of the two she can safely trust. Indeed, making hateful citizens visible helps targets of hateful views trust trustworthy citizens *and only* trustworthy citizens.

In sum, the objection that not excluding hate speech is bound to undermine trust is too hasty. By making hateful citizens more visible, the practice of allowing hate speech while authoritatively denouncing it performs two trust-building functions: it creates new opportunities for persuading hateful citizens; and when persuasion fails, it at least

⁷⁵ Schulzke (2016, 230-236).

⁷⁶ Cited in Chakravarti (2014, 69).

⁷⁷ Schulzke (2016, 230-235).

makes it clearer who can and cannot safely be trusted. This second point concedes that hate speakers and their targets will not trust each other. Even so, I have suggested that the fundamental trust-related problem here is the *existence* of hateful citizens. Their untrustworthiness with respect to their targets is what destabilises trusting relations, rather than the policy of allowing those hateful citizens to air their views. Thus, the more crucial point against premise (3) remains: *given* that hateful citizens exist, the policy of allowing public hate speech (while authoritatively denouncing it) arguably seems to be part of the solution, rather than the source of the problem.

5. Conclusion

According to the problem of goodwill, inclusive deliberation presupposes unrealistic levels of mutual goodwill and (as a presupposition thereof) trust. Moreover, it is said that deliberation cannot itself contribute to generating trust and the attending goodwill in non-ideal conditions.

The present chapter has articulated reasons for pushing back against this challenge. To begin, I showed that on the systemic understanding of deliberation, discursive processes presuppose less goodwill than one might think: highly demanding discursive norms needn't apply to most deliberative spheres; and when the norms which do apply to most citizens—e.g., norms of sincere engagement—remain too demanding, the networked interaction between different spheres helps circumvent the communicative barriers raised by distrust.

These qualifications notwithstanding, the challenge remains problematic. Accordingly, most of my argument tackled another claim: that the public deliberation which characterises non-ideal contexts cannot generate greater trust and goodwill. Specifically, I explored respects in which three features of such public discourse—hypocrisy, angry speech, and the occurrence of hate speech—offer resources for rebuilding trust and the attending goodwill.

Even if we start from pessimistic assumptions regarding existing goodwill and trust, the practice of hypocritically pretending to be public-spirited and sincere may, as Elster suggests, make participants more accountable to their interlocutors. Public hypocrisy induces participants to listen, and eventually to answer challenges and to change their stated positions.

Given a disposition to listen, other characteristic features of non-ideal public discourse have promising trust-building properties. Drawing on philosophical approaches to trust, anger, and transitional justice, I theorised several trust-building functions of personal narratives charged with anger. Telling one's anger-infused story can have a humanising influence on the listener's perception of the speaker, signal trust in a way that invites the listener's trustworthiness, help the speaker feel acknowledged, and provide information that is instrumental to rebuilding social bonds.

But one problem with making space for angry testimony in divided contexts is that testimonies expressing unfitting anger might conceivably veer into hate speech. Nevertheless, *given* a situation where some individuals have hateful views, it is unclear that allowing public hate speech while authoritatively denouncing it makes matters worse for trust. Instead, as Schulzke contends, there are reasons to think that it improves social trust by making it clearer who can and cannot be trusted.

Notice, finally, how this response to the problem of goodwill accords with the normative recommendations of previous chapters. Emotionally-charged narratives, including angry testimony, are forms of speech I defended in Chapters 2 and 4. And Chapter 3 recommended legally allowing public hate speech while having authoritative public figures denounce it. A similar observation holds with public hypocrisy. The occurrence of fruitful hypocrisy results from the expectation that participants in public discourse be sincere—a norm advanced in Chapter 1, Section 3.1—as well as by the expectation that, in some spheres, participants should appeal to considerations that are public or shared—a norm defended in Chapter 2, Section 2.2.

I originally defended these norms on the grounds that they play a key role in realising justice. But the fact that these very same norms offer *prima facie* promising resources for generating trust and goodwill in divided contexts should bolster this initial defence. It suggests that this normative account of democratic public deliberation is well-suited to withstanding the shortage of goodwill which characterises divided political settings.

CHAPTER 6

DEFLECTING THE PROBLEM OF PUBLIC IGNORANCE

1. Introduction

Citizens of contemporary democracies are often extremely ignorant about political matters. Many are disinclined to seek out and accept information from trustworthy sources. The extent of this ignorance can be startling. American citizens commonly do not know which party controls Congress. Similarly, there is evidence that most citizens cannot name the three branches of federal government, let alone explain their function. And even during the Cold War, only a small minority of Americans knew that the USSR did not belong to NATO.¹

This ignorance constitutes a pressing political problem. In brief, ignorance about public affairs—public ignorance—makes citizens vulnerable to demagogues, who can mislead them into supporting unjust policies or unqualified politicians. This danger was arguably on display in 2016 when, some analysts have suggested, ignorance about the European Union² and about American politics³ contributed to the UK's decision to leave the EU and to the election of Donald Trump.

Because public ignorance seems politically dangerous, political philosophers have become increasingly interested in how we might mitigate its negative impact. In this context, an influential objection holds that normative theories of democracy that give a central role to democratic public deliberation—to public deliberation involving ordinary citizens on a large scale—are uniquely ill-adapted to the fact of public ignorance. Fruitful deliberation, it is said, requires citizens to be more knowledgeable than they actually are about politics. This shortcoming is standardly taken to warrant adopting alternative normative models of democracy, or even non-democratic political systems.

The present chapter will examine this pressing objection and its effectiveness against the normative account of democratic deliberation I have advanced. Thus, while

¹ Somin (2013, 18-22).

² Somin (2016b).

³ Somin (2016a).

Chapter 5 considered the objection that my framework presupposes unrealistic levels of goodwill and trust between participants, the present chapter will investigate the objection that it assumes unrealistic dispositions to acquire political knowledge.

Though public ignorance constitutes a serious concern, I will argue that it is not a special problem for deliberative theories of democracy like my own. Exploring the alternatives to deliberation-centred democracy reveals *either* that their recommendations are equally plagued by public ignorance, *or* that, although their recommendations constitute promising remedies for public ignorance, they can be reconciled with deliberative democracy.

After introducing the difficulties public ignorance allegedly generates for inclusive deliberation (Section 2), I briefly consider the response that ignorance does not actually impede deliberation's ability to promote justice. For reasons that are partly dialectical, I concede that this response is unsatisfactory (Section 3). Instead, I demonstrate that if public ignorance is problematic for deliberative democracy, then it also undermines alternative arrangements. First, more minimalistic democratic arrangements—which either eschew deliberation by ordinary citizens or restrict it to small-scale settings—neither avoid nor alleviate the problem, despite demanding less active participation from citizens. And insofar as these models *do* seem less vulnerable to ignorance, this is because of features that are ultimately consistent with deliberative democracy (Section 4). Now, if all models of democracy are vulnerable to the present problem, perhaps the solution is to abandon democracy. However, I will show that the non-democratic arrangements touted as remedies in fact merely postpone the problem of public ignorance (Section 5).

The point of this companions-in-guilt argument is not solely negative. Its positive purpose is to refocus normative theorising about democracy. First, it reveals that criticisms that move from public ignorance to the rejection of deliberative democracy are distracting. If alternative normative models are also plagued by widespread ignorance, our focus should be not on departing from deliberation-centred democracy, but rather on tackling the problem at its source: public ignorance itself.

Second, to combat public ignorance, we should at least partly direct our attention to improving practices of mass public deliberation. Since everyone faces the problem of public ignorance, we should think of inclusive deliberation not as the source of the problem, but rather as an attempted solution to it. Given that the problem is

ubiquitous, and given the promise of contemporary deliberative innovations, conceptual and institutional, we must try to make this solution work.

2. The Problem of Public Ignorance

2.1. The Fact of Public Ignorance

Empirical research indicates that most citizens are ignorant regarding various dimensions of politics. Their ignorance concerns policy outcomes (what the outcomes of given policies might be), policy availability (what policies are available), policy support (which political actors support which policies, and what they are doing about it), and the policy-making process (what role different political actors have in developing and enacting policies).⁴

A major cause of this ignorance is that citizens typically lack the disposition to seek out and accept reliable political information. Some citizens hardly look for information at all. Others seek information, but only from a very restricted range of sources. And many citizens search for information but trust unreliable sources. These features are interrelated. Knowing very little and seeking information from few sources makes it more likely that one will end up trusting unreliable sources, whose distorted claims lack independent support. In turn, trusting unreliable sources may erode one's disposition to consider other, perhaps more reliable, sources.⁵

What is more, these dispositions may be difficult to alter. When there are many unreliable sources, and when people with different perspectives are segregated into different social enclaves, gathering reliable and balanced information is costly. At the same time, the probability of making a difference with one's vote is minimal. Hence, some social scientists argue that being reluctant to search for high-quality information is rational.⁶ So, ignorance regarding public matters is widespread and originates in dispositions that are not only mutually reinforcing but may even be instrumentally rational.

⁴ For overviews, see Somin (2013, ch. 1) and Brennan (2016, ch. 2).

⁵ For discussion, see Anderson (2011, 153-154), Parkinson (2006, 103-108), and Anderson (1998, 498-500).

⁶ E.g., Somin (1998, 435-438).

2.2. The Threat to Inclusive Deliberation

This public ignorance, it is often argued, is especially problematic for conceptions of democracy that advocate inclusive deliberation. At this point, it is worth reiterating a few notable features of what deliberative theories of democracy do and do not say. First, although deliberative democracy morally requires active participation by ordinary citizens on a large scale, it is not synonymous with direct democracy. As discussed in Chapter 1, deliberative democrats usually take inclusive deliberation to be intimately bound up with political representation. That is, a significant part of what citizens are expected to deliberate about is their actual or potential representatives: in a representative deliberative democracy, citizens debate which representatives they should elect; they argue about whether their elected representatives are enacting good policies; they write to those representatives to demand justifications; and they publicly criticise representatives who break electoral promises.⁷ Deliberating about one's representatives will inevitably involve arguing about the merits of different policies—for instance, when assessing competing politicians' policy platforms. But it can also involve discussing other issues, such as politicians' character.

Relatedly, large-scale deliberation does not require an immense deliberative arena that includes all citizens. As emphasised in previous chapters, we should understand large-scale deliberation as occurring within a deliberative system—i.e., a network composed of many deliberative arenas, which vary in size, formality and function, and are connected more or less closely to one another.⁸ While formal deliberative arenas—such as legislative assemblies—are often tasked with producing coercive decisions, more informal arenas—like newspapers' opinion sections or political rallies—principally aim to help citizens develop more considered views.

The most essential feature of deliberative democracy, then, is the central role it gives to public deliberation that includes ordinary citizens on a large scale. Why is ignorance problematic for such deliberation? The objection can be reconstructed as follows:

⁷ Chapter 1, Section 3.1.

⁸ Chapters 2, 4, 5.

- (1) Public deliberation promotes just outcomes only if participants are well-informed.
- (2) Ordinary citizens are typically not well-informed.
- (3) So, inclusive public deliberation is typically not justice-promoting.⁹

Premise (2), as we have seen, is strongly supported by empirical research. As for premise (1), it seems *prima facie* plausible. In earlier chapters, I argued that inclusive public deliberation contributes to promoting justice on epistemic and motivational grounds. On the epistemic side, exchanging arguments and testimonies, both dispassionate and emotionally-charged, helps interlocutors discover which policies are most just and which potential representatives are most likely to promote justice.¹⁰ On the motivational side, publicly questioning powerful actors helps to hold them accountable. Thus, public deliberation constitutes a check on powerful agents, which prevents them from exercising their power arbitrarily and thereby dominating weaker parties.¹¹

The problem is that both mechanisms seem to require that participants satisfy some threshold of knowledge. If A is highly ignorant, A might fail to notice that B's arguments rest on false assumptions, that C's emotional testimony is fictitious, or that D's current assertions are inconsistent with her earlier assertions. In turn, if A cannot distinguish good arguments and testimonies from bad ones, the epistemic and motivational benefits of deliberation may disappear. Indeed, if A is equally swayed by spurious and non-spurious arguments and testimonies, deliberation might not make him more knowledgeable concerning which policies and politicians are better at advancing justice. Likewise, if A cannot determine whether his representatives' claims are false or inconsistent, then publicly questioning them will be ineffective at holding them accountable. Hence, standard defences of inclusive deliberation (including the ones I have been appealing to) might seem to depend upon unrealistic assumptions regarding participants' knowledge.

For many theorists, this shortcoming warrants shifting to normative models of government that demand less of ordinary citizens than inclusive deliberation does—and, accordingly, that seem more sensitive to the fact of public ignorance. While some

⁹ E.g., Anderson (1998, 490-491), Somin (1998, 438-442), Posner (2003, 163).

¹⁰ Chapter 1, Section 3.1; Chapter 2, Section 4; Chapter 4, Section 4.

¹¹ Chapter 1, Section 3.1; Chapter 2, Section 2.2; Chapter 5, Section 4.1.

advocate more minimalistic theories of democracy, which remove deliberation by ordinary citizens or restrict it to small-scale settings, others defend non-democratic political systems. In what follows, I briefly examine attempts at rejecting the claim that ignorance undercuts the justice-promoting powers of inclusive deliberation (Section 3). My focus, however, will be on demonstrating that even if we accept this claim, it does not follow that we should embrace an alternative normative model (Sections 4-5).

3. Is Ignorance Really a Problem for Democratic Deliberation?

Defenders of deliberation-centred democracy often object to premise (1). Put differently, they argue that even if citizens are *initially* highly ignorant about politics, this is not a deep problem for deliberative democracy's capacity to advance justice. On this line of argument, inclusive communication is a remedy for ignorance, which makes participants less ignorant than they initially were. Consequently, even a deliberative process that takes as its input many ignorant citizens can ultimately produce a wise and justice-promoting output.¹²

Consider two ways in which deliberation might reduce ignorance. First, when some citizens—call them ‘experts’—know substantially more than others about some political issue, they can share this knowledge with ignorant co-citizens. Ignorant citizens can learn *that policy x is good (bad), that claim y is true (false)*, etc., by consulting and questioning experts on political matters. This is a familiar feature of actual public communication. Citizens regularly turn to fact checkers, economists, media outlets, and opinion leaders more generally for information.¹³

Second, even when no participant is an expert concerning the issue at hand, the relevant political knowledge might still be distributed between the participants. Thus, many individually ignorant citizens might pool the knowledge they have by deliberating together. While this claim is commonly gestured at by deliberative democrats, Hélène Landemore has developed it most fully. Drawing on formal mathematical models such as the Hong-Page theorem, she argues that under the right circumstances, a diverse

¹² E.g., Landemore (2013, 50).

¹³ On deliberation with experts, see Anderson (2011, 144) and Christiano (2012).

group of non-experts can put their cognitive resources together, and thereby produce even wiser decisions than a group of experts.¹⁴

However, these two ways of challenging premise (1) encounter similar difficulties. Take the latter. This strategy seems to idealise away the social conditions that give rise to the problem of public ignorance. Landemore holds that the exchange of arguments and information between diverse non-experts can reliably make them more knowledgeable. But this is only true if the non-expert participants already meet a threshold of knowledge. As discussed in 2.2, when participants cannot discriminate between good and bad arguments and testimonies, deliberation can lead them to pool their ignorance as well as their knowledge. Hence, as many of Landemore's critics complain, her answer to the problem of public ignorance seems to presuppose the falsehood of one of its premises—namely, that people are highly ignorant to begin with.¹⁵ In fact, when she states the conditions that must obtain for deliberation to be epistemically fruitful, Landemore requires that participants be sufficiently knowledgeable to recognise the best policy when they see it.¹⁶

The former strategy—which involves consulting expert testimony—might seem less vulnerable to this concern. This is because the listeners are not trying to critically assess the arguments and testimonies on the basis of which experts arrive at judgments. They simply take their claims on trust.

But this strategy quickly runs into problems. The first is dialectical. Chapter 5 investigated how democratic public deliberation can build trust. One of the main trust-building mechanisms I appealed to—the civilising force of hypocrisy—presupposes that individuals have decent political knowledge, which enables them to monitor public speakers. In this context, it would be viciously circular to appeal to pre-existing relations of trust—namely, the disposition to trust others' testimony—to explain how political knowledge arises.

Even if we bracket these dialectical considerations and assume that individuals are disposed to trust others' testimony, the deeper problem is that public ignorance makes it more difficult for them to engage in *warranted* relations of trust—i.e., to trust the testimony of *trustworthy* speakers. Deferring to A's testimony that P is the best policy means that one needn't examine competing arguments regarding the intrinsic merits of

¹⁴ Landemore (2013, ch. 4). See also note 44, Chapter 1.

¹⁵ E.g., Somin (2014, 160), Brennan (2016, 187-188), Gunn (2014, 68-69).

¹⁶ Landemore (2014, 218-221).

P. Nevertheless, one still needs to debate and examine competing arguments concerning whether or not A is actually a trustworthy source, a real and honest expert. Competently doing so requires substantial information, for instance regarding which credentials A has, whether A has conflicts of interest, etc. Therefore, in conditions of public ignorance, individuals may well end up dismissing the testimony of real experts, and trusting the testimony of mere sophists instead. Even Elizabeth Anderson, one of the more optimistic theorists of reliance on expert testimony, concludes that “the dispositions to reliable assessment [of experts] depend on social conditions that are not realized in the U.S.”¹⁷ Like the other strategy, then, the idea of consulting experts seems to disregard the central premise of the problem of public ignorance: that citizens are typically highly ignorant.

This objection might seem too hasty. Landemore, for instance, could reply that it misconstrues her point. The point is not necessarily that individuals are not ignorant. It is rather that even if they are, they still meet the threshold for deliberation to be epistemically fruitful. Put differently, critics of deliberative democracy are wrong about how high the knowledge threshold for deliberation to be effective is. Of course, it is very difficult to give a specific account of where the threshold lies. But doing so, Landemore suggests, is not necessary. If deliberation is in fact epistemically fruitful, then *a fortiori*, its participants—though ignorant—satisfy the threshold.¹⁸ And according to Landemore, there is ample evidence of the epistemic benefits of deliberation by ordinary citizens. Famously, when randomly-selected citizens engage in structured deliberation on a policy issue—as in Citizens’ Assemblies on electoral and constitutional reform, and so-called “deliberative polls” examining competing energy policies—their participants tend to become far more knowledgeable about the relevant issue.¹⁹ So, even if ordinary citizens are ignorant, empirical evidence suggests that this ignorance is insufficient to prevent them from learning through deliberation with peers and experts.

However, it is unclear how generalisable this response is. Landemore’s evidence typically concerns deliberation occurring in highly controlled settings: the deliberative experiments in question involve small groups; their deliberation is heavily facilitated by

¹⁷ Anderson (2011, 145-149).

¹⁸ Landemore (2014, 207).

¹⁹ *Ibid.*, 208-211. See also Landemore (2013, 109-110).

moderators and experts; and the participants often share certain aims or goals.²⁰ Therefore, Landemore's defence of the epistemic benefits of deliberation tells us little about what to expect from deliberation between ignorant citizens in the more familiar arenas of mass democracy, many of which are greater in scale, less structured, and marked by significant ethical divisions. I will return to these controlled instances of deliberation in Sections 4 and 5. For now, the point is that they do not warrant generalisations about large-scale deliberation between ordinary citizens.

What is worse, there is evidence that in some non-ideal circumstances, large-scale deliberation does pool ignorance as well as knowledge. Notably, Cass Sunstein observes a "law of group polarisation". When like-minded groups deliberate, their views tend to shift towards more extreme versions of their original judgments. In part, this is because like-minded groups have access to an imbalanced pool of arguments: they mostly produce arguments supporting the position they already hold, and have little exposure to countervailing perspectives.²¹ While this shift towards more extreme positions can in principle lead to truer views, it can equally lead to falser ones. A striking example concerns online deliberation during the 2016 US Presidential election. Social media sites, like Facebook, allow news to spread extremely rapidly among like-minded individuals, in a way that is insulated from contestation by fact-checkers and from dissenting voices. In consequence, during the election, false news stories concerning the two candidates "were more widely shared on Facebook than the most popular mainstream news stories". Many of those exposed to these false stories reported believing them.²² In this instance, deliberation in large-scale online enclaves helped circulate and amplify erroneous views rather than eliminate them.

Perhaps some institutional remedies can improve the ignorance-reducing prospects of exchanging arguments with epistemic peers and of questioning purported experts in non-ideal conditions. For instance, Elizabeth Anderson recommends institutional measures aimed at guaranteeing that citizens are exposed to diverse perspectives. These include reorganising political districts and forcing news sites to include links to sites advocating opposite views. Anderson hypothesises that such measures will prevent groups from polarising in a truth-insensitive way, and will help deliberation produce

²⁰ For criticisms of how atypical Landemore's examples of deliberation are, see Gunn (2014, 68) and Brennan (2016, 66-67).

²¹ Sunstein (2002).

²² Alcott and Gentskow (manuscript, 1-2).

better-informed views regarding which policies are best and which experts are trustworthy.²³ However, she ultimately concedes that “discussion of remedies is hampered by the lack of empirical data”.²⁴

For dialectical reasons, then, the rest of this chapter will not rely on the current line of response to the problem of public ignorance. Indeed, I do not want my response to the problem of public ignorance to rest solely on the disputed claim that deliberation between highly ignorant citizens remains epistemically fruitful, or that feasible institutional remedies can readily make it so. If these claims can be vindicated, then so much the better: my arguments in previous chapters regarding the justice-promoting power of deliberation, many of which explore its ability to reduce ignorance, go through. But in the meantime, I wish to examine what could be said in defence of deliberative democracy even if ignorance *is* highly problematic. Therefore, I will argue that even if ignorance undermines the fruitfulness of democratic deliberation, this is no reason to abandon it in favour of an alternative normative political model. In doing so, my aim is indirectly to support the response just considered: if adopting alternative political arrangements were a solution to the problem at hand, we might lose patience with ongoing conceptual and institutional attempts at making inclusive deliberation capable of reducing ignorance. However, if doing so is no solution, as I will demonstrate, we have little choice other than to try to make these attempts work.

4. Why Ignorance Is a Problem for Alternative Models of Democracy

If public ignorance undermines democratic theories that require large-scale inclusive deliberation, one might recommend adopting a conception of democracy that is more sensitive to the limitations of ordinary citizens. Against this influential argument, the present section argues that the two most promising democratic models that have been advanced to circumvent the problem of public ignorance fail to do so.

²³ Anderson (2011, 157-160).

²⁴ Ibid., 159.

4.1. Elite Democracy and Ignorance

To avoid the problems public ignorance poses for inclusive deliberation, many commentators advocate a minimalistic model of democracy—elite democracy—in which ordinary citizens play a smaller part. Elite democracy is a resolutely representative form of democracy. Citizens do not vote on every policy decision. They merely elect more competent political representatives—elites—and throw them out in subsequent elections when they perform poorly. Elite democracy therefore staunchly opposes direct democracy. But this is not what distinguishes it from deliberative democracy. As discussed in 2.2, deliberative democrats typically embrace political representation. Instead, elite democracy’s central difference with deliberative democracy is that ordinary citizens are not morally required to engage in public reason-giving with each other and with their representatives. The only essential democratic act is voting for representatives.²⁵

This elite theory of democracy has most famously been championed by Joseph Schumpeter and neo-Schumpeterians such as Richard Anderson, Richard Posner, Jeffrey Green, and Ian Shapiro.²⁶ On the Schumpeterian account, self-interested elites compete for office. Like firms targeting consumers, they develop packages of policies and attempt to ‘sell’ them to voters. Citizens then vote for the elites who seem most likely to address their concerns, shunning elites whose policies seem undesirable and replacing those who have proven incompetent in power. The competition for office is intended to guarantee that elites will offer adequate policy packages. Likewise, the threat of being voted out of office motivates elected elites to keep their promises.

Besides this minimal role, however, citizens should be largely uninvolved in politics. According to Schumpeter, citizens “must understand that, once they have elected an individual, political action is his business and not theirs”.²⁷ This passivity extends to deliberation. Richard Anderson, for instance, claims that citizens are not morally required to deliberate.²⁸ But elite democrats generally go further. Schumpeter positively

²⁵ Elite democracy’s rejection of direct democracy *does* distinguish it from another non-deliberative theory of democracy: aggregative democracy. Like elite democracy, aggregative democracy emphasises the aggregation of votes over inclusive deliberation. But aggregative democracy is consistent with direct democracy, where citizens vote on all issues.

²⁶ Schumpeter (1942); Anderson (1998); Posner (2003); Green (2010); Shapiro (2016, 78-83).

²⁷ Schumpeter (1942, 262).

²⁸ Anderson (1998, 494).

demands that citizens refrain from writing to their representatives.²⁹ And, perhaps most strikingly, Green affirms that democracy should be “ocular”: the role of ordinary citizens is “characterized by *silence* rather than decision, *spectatorship* rather than activism”.³⁰

Schumpeter, Posner, and Anderson all take public ignorance as a key motivation for adopting elite democracy. And although they do not explicitly discuss ignorance, Shapiro and Green prefer elite democracy because they find it more realistic than deliberative accounts.³¹ There is something intuitively appealing about this. *Prima facie*, if ordinary citizens are so ignorant that deliberation is as likely to be damaging as it is to be helpful, it makes sense to adopt a model of democracy that is sensitive to these limitations. By jettisoning deliberation, and restricting citizens’ role to voting out bad leaders, elite democracy does this more emphatically than any other conception of democracy.

But this appeal is an illusion. Public ignorance also constitutes a grave problem for elite democracy. Although elites are responsible for the overwhelming majority of day-to-day politics, there remains a bottleneck of democratic decision-making—elections—where ordinary citizens have the final say. Making just decisions during elections requires various kinds of political knowledge. To compare competing politicians’ policy platforms, one must know which policies are on offer (policy availability), which politicians support which policies (policy support), and what the likely outcomes of different policies are (policy outcome). Moreover, to hold politicians accountable, one must know who has been responsible for doing what in the policy-making process (policy-making process). These, as discussed in 2.1, are precisely the kinds of knowledge that ordinary citizens are said to lack. Consequently, when individuals are ignorant, elite democracy is problematic: there is no guarantee that politicians will offer proposals that genuinely cater to citizens’ interests, no guarantee that citizens will choose the best policy platforms among those offered, and no guarantee that politicians will keep their promises.

One might nevertheless maintain that public ignorance is a *greater* problem for deliberative democracy than for elite democracy. For Somin, this is because theories of

²⁹ Schumpeter (1942, 262).

³⁰ Green (2010, 189).

³¹ Schumpeter (1942, 234); Posner (2003, 151-152); Anderson (1998, 493-499); Green (2010, 24); Shapiro (2016, 6-9).

“deliberative democracy [...] require greater knowledge than less onerous ones such as Schumpeterian retrospective voting”.³² First, while both elite and deliberative democracy demand knowledge of policy outcomes, policy availability, policy support, and of the policy-making process, “deliberative democracy probably requires greater factual knowledge of this kind, because voters are expected to engage in *serious* deliberation on policy alternatives”.³³ Second, because of the norms that regulate deliberation, deliberative democracy also requires further kinds of knowledge, “moral and philosophical knowledge”, which elite democracy does not require. Citing Gutmann and Thompson, Rawls, and Habermas, Somin claims that deliberation involves appealing to mutually acceptable or shared considerations.³⁴ Hence, participants must know not only which vote would promote desirable outcomes, but also which considerations are shared.

Neither of these two grounds for thinking that deliberative democracy demands more knowledge is compelling. First, it is unclear why deliberators must know more about policy availability, support, outcome, and process than voters in elite democracy. The normative appeal of elite democracy depends crucially on voters being able to identify bad representatives and replace them with better elites. Thus, insofar as elite democracy is normatively appealing, it too requires that voters reflect seriously on competing politicians and their policy platforms. Put differently, even if voters in elite democracy do not need to engage in serious interpersonal deliberation, they still need to engage in serious *intrapersonal* deliberation if they are to vote judiciously. And there is no reason to think that serious intrapersonal deliberation about politicians and their platforms demands less factual knowledge than serious interpersonal deliberation about these issues.

The second suggestion—that democratic deliberation requires further kinds of knowledge—is more promising, but ultimately misses its mark against many deliberative theories of democracy, including my own. Chapter 2 observed that the norm according to which participants must appeal to shared reasons should only apply in formal deliberative sites, such as the legislature. Accordingly, the vast majority of citizens needn’t comply with it, and needn’t possess the corresponding knowledge. Now, some other kind of moral or philosophical knowledge may be required of

³² Somin (2014, 160).

³³ Somin (2013, 43), emphasis added.

³⁴ Ibid., 42-43.

ordinary deliberators. Deliberators are generally expected to argue for the *moral* desirability of policies, which requires knowing about moral values. But this does not seem specific to democratic deliberation. To vote for elites in a justice-promoting way, one must know not only what the causal effects of one's vote might be, but also how just those effects are. That simply is moral knowledge.

Furthermore, even if fruitful deliberation *did* require more information than competently replacing elites, it still would not necessarily follow that public ignorance is a greater problem for deliberative democracy. Suppose we are considering two strategies, S_1 and S_2 , for achieving a goal G . Both strategies require us to have information that happens to be unavailable. But S_1 requires less unavailable information than S_2 . Even then, if the unavailable information that S_1 requires is necessary for S_1 to achieve G , and if that information is genuinely inaccessible, then S_1 may be just as unsuccessful at achieving G as S_2 . To see this, imagine that we are trying to open a door, and two passcodes are needed to do so. A only knows one of the passcodes, and B knows neither. In this case, asking A to open the door will be just as unsuccessful as asking B, even though the strategy of asking A requires less unavailable information than that of asking B. Thus, from the fact that a strategy S_1 requires less unavailable information than S_2 , it needn't follow that ignorance is a greater problem for S_2 than for S_1 . Consequently, even if elite democrats established that deliberative democracy requires that ordinary citizens be more knowledgeable than elite democracy does, this would not necessarily settle the matter. In conditions of political ignorance, both might conceivably be equally ineffective at promoting justice.

At this stage, elite democrats might take a different line of argument, and object that my characterisation of elite democracy is uncharitable. I have suggested that in elite democracy, elites simply present their policy platforms, before letting citizens choose. But this is arguably not all there is to elite democracy. Instead, Green claims, elites should also offer arguments for their policies. Notice that this is elite deliberation, not democratic deliberation: elites exchange arguments, while ordinary citizens are primarily tasked with listening.³⁵ The idea is that elites' arguments for their proposals will educate ordinary citizens about politics. Hence, citizens will be more knowledgeable when they elect elites.

³⁵ Green (2010, 182-186). See above, including notes 28-30, for the lack of deliberation by ordinary citizens.

Deliberative democrats often object that such elite-driven discourse is deficient. First, elites may have substantial epistemic gaps. Indeed, as Chapter 1 discussed, politically-relevant knowledge is distributed between different social groups, and some social groups are underrepresented among elites.³⁶ Second, even if elites had all the politically-relevant knowledge, Simone Chambers fears that they may be ill-motivated. They might collude to keep some ideas off of the agenda, while offering deceptive arguments aimed at making people accept policies that only serve elite interests.³⁷

But let us assume that some elites know what the most just policies are, are motivated to advance them, and offer arguments for them. The main problem is this: why think that ignorant voters will be swayed by the arguments of wise and benevolent elites, rather than by the sophisticated appeals of unwise or self-interested demagogues? The point is that even if there are real elites, who are knowledgeable and well-motivated, not all candidates for office are like this. Therefore, for elite deliberation to educate listeners, ordinary citizens would need to discriminate between the sound appeals of real elites and the deceptive ones of mere sophists. If elite democrats respond that voters know enough to tell who the trustworthy elites are, or to distinguish good from bad policy arguments, they are helping themselves to an assumption that they denied deliberative democrats, in Section 3. So, appealing to elite deliberation to defend elite democracy ends up making epistemic demands on ordinary citizens that parallel those of deliberative democracy.

As a final response, elite democrats might suggest that candidates for office are actually almost invariably more knowledgeable than the average citizen, and are often decently motivated. And even if some are not, they might add, anyone who is elected is *then* made to learn about politics and to behave accountably by their aides and colleagues. If this is the case, ordinary citizens' ignorance about politics might seem relatively unproblematic for elite democracy. Because all possible representatives would end up being politically competent, the stakes of elections are low. No matter how ignorant voters are, they cannot go too far wrong.

However, this response would not give elite democracy any comparative advantage: if we grant its premises, deliberative democrats too can help themselves to this defence. The response under consideration is really an argument for representative democracy: it

³⁶ Chapter 1, Section 2.1.b.

³⁷ Chambers (2009, 338-339).

says that, in conditions of political ignorance, having citizens elect representatives is better than having them make policy directly. But, as discussed in 2.2, deliberative democracy is entirely consistent with representative democracy. Accordingly, if all representatives really are or would become competent, deliberative democrats might say the following: ‘Even if deliberation between ignorant citizens about candidates *were* fruitless—i.e., even if citizens tended to ask candidates the wrong questions and to be swayed by bad answers—this would not be exceedingly problematic: since all of the candidates would ultimately be competent in office, deliberative participants cannot go too far wrong when they vote’.³⁸ Thus, if representative democracy mitigates the negative impact of public ignorance, this helps deliberative democracy as much as elite democracy.

Elite democracy responds to the ignorance of ordinary citizens by getting them to do less. However, it retains a bottleneck of democratic decision-making, elections, where there is no reason to think that ignorant voters will choose the correct elites—i.e., those who are both knowledgeable and well-motivated. Consequently, I have argued, widespread ignorance about politics is just as problematic for elite democracy as for deliberative democracy. If, to resist this objection, elite democrats underscore the educational benefits of elite deliberation, they replicate the informational burdens of deliberative democracy. And if, instead, elite democrats appeal to the alleged competence of representatives, this response is equally available to deliberative democrats.

³⁸ This reply might encourage the worry that deliberative democracy is redundant, because inclusive deliberation itself plays no part in avoiding the problems raised by public ignorance. This worry is misplaced for two reasons. Firstly, the present reply does not entail that inclusive deliberation plays no part in countering public ignorance. The only essential claim I am making here is that *if* you think political representation makes widespread political ignorance unproblematic, this is just as true for deliberative as for elite democracy. Secondly, even if inclusive deliberation did not help reduce ignorance when citizens are highly ignorant (as I conceded for dialectical reasons at the end of Section 3), this would not mean that deliberative democracy is altogether redundant or unnecessary. All it would mean is that deliberative democracy has no advantage over elite democracy *with respect to the problem of public ignorance*. Let me clarify what I mean by this. In previous chapters, I argued that inclusive deliberation can perform important justice-promoting functions, such as creating knowledge or preventing domination. The problem of public ignorance says that for deliberation to perform these functions, deliberators must satisfy some prior threshold of knowledge. *When they do not*, deliberative democracy may admittedly have no comparative advantage over elite democracy. This is fully consistent with the present chapter’s central argument: I am arguing that, in such conditions, elite democracy also has no advantage over deliberative democracy, so that public ignorance gives us no reason to abandon deliberative democracy for elite democracy. Crucially, however, *after the knowledge threshold is satisfied*, deliberation’s distinctive justice-promoting functions kick in, and these give us a reason to prefer deliberative democracy. For discussion of this dialectic, see the thesis Conclusion, Section 1.

Thus, elite democracy fails to circumvent the difficulties raised by widespread ignorance, and reemphasises the significance of deliberation when it tries to address this issue. In this light, inclusive deliberation aimed at making ordinary citizens more knowledgeable should be seen not as the source of democracy's ignorance-related difficulties, but rather as an attempted solution to these. Perhaps, then, the proper response to public ignorance is not to eliminate deliberation by ordinary citizens—as elite democracy does—but rather to restrict it to more structured settings. The next section examines this popular suggestion.

4.2. Is Small-Scale Deliberation the Answer?

To evade elite democracy's ignorance-related shortcomings, a prominent solution recommends reinstituting deliberation by ordinary citizens, but doing so in small-scale settings, or “minipublics”. According to James Fishkin and Robert Luskin, who have most prominently advanced this response, the fact that most citizens are ignorant and do not engage in high-quality deliberation is “unlikely to change very much. Not for the public as a whole, at least”.³⁹ But it might change for a smaller subset of the public.

This proposal aims to achieve the best of both worlds. Since, as 4.1 discussed, elite democracy is plagued by the ignorance of ordinary citizens, it aims to educate samples of ordinary citizens through small-scale deliberation. At the same time, it promises to avoid the problems raised in Section 3 with using large-scale deliberation to educate ordinary citizens. Small-scale deliberation, it is said, can be structured so as to enable ordinary citizens to learn effectively. Thus, small-scale deliberative settings avoid the chaotic mass public sphere's limitations.

These claims are not simply speculative. Political scientists have extensively experimented, with promising results, on the use of minipublics—small representative samples of the population that deliberate on specific issues—in policy-making. Section 3 briefly alluded to some of these. In Fishkin's early “deliberative polls”, small groups of randomly-selected citizens in Texas deliberated on the merits of investing in energy efficiency and renewable resources. In 2004, a Citizens' Assembly composed of 160 randomly-selected citizens was tasked with deliberating regarding which electoral system British Columbia should adopt. Most impressively, between 2010 and 2011,

³⁹ Fishkin and Luskin (2005, 286). See also Parvin (2015, 419) and López-Guerra (2014).

Iceland charged groups of randomly-selected citizens with designing, via deliberative processes, a new national constitution.⁴⁰

Consider how these minipublics realise the goals outlined above. First, participants are selected randomly. The resulting groups are therefore likely to constitute a representative sample or ‘mirror’ of the population. So, unlike elite democracy, minipublics actively include ordinary citizens in deliberation. Second, the groups are sufficiently small (between 20 and 500) for their deliberation to be carefully structured. Generally, deliberation is facilitated by trained moderators, who ensure that all participants are heard. Furthermore, deliberation is interwoven with briefings by experts, whom participants can question. Finally, the epistemic results are impressive: Fishkin and Luskin observe that participants habitually become significantly more knowledgeable regarding the issues at hand, and that their policy preferences often change accordingly.

But the proposal of introducing small-scale structured deliberation by representative samples of the population is ambiguous between two possibilities. On one interpretation, minipublics make political decisions themselves, thereby *replacing* existing decision-making arenas. On another interpretation, minipublics play an advisory role, providing information and recommendations to the broader demos and their representatives. Hence, they *complement* existing processes of mass decision-making, as experts might.

Most democratic theorists who advocate using minipublics adopt the latter understanding.⁴¹ This is mirrored in practice. A key function of deliberative polls, for instance, is to inform broader public debates. Accordingly, media outlets often broadcast their proceedings, thereby exposing the people and their representatives to the views of citizens who have reflected on specific policy issues, and to the arguments underpinning those views. Even the British Columbia Citizens’ Assembly, which was one the most empowered minipublics, only had the power to recommend an electoral system, which the people then needed to approve by referendum.

If this is the role of minipublics, then this proposal is entirely consistent with the kind of inclusive deliberation I have been defending. Recall that on the systemic understanding of deliberative democracy, the deliberative system is composed of many

⁴⁰ Goodin and Dryzek (2006); Fishkin and Luskin (2005); Landemore (2015).

⁴¹ E.g., Goodin and Dryzek (2006, 220), Fishkin and Luskin (2005).

interconnected arenas of deliberation, which perform complementary roles. Considered in this light, advisory minipublics are just one arena in the deliberative system, whose purpose is to transmit information to more empowered arenas.⁴²

Unsurprisingly, then, the problem that afflicted mass democratic deliberation in Section 3 resurfaces. If minipublics are just one arena among others, and their purpose is to convey information to empowered spheres, why think that the broader public sphere will heed their advice? Just as ignorant citizens may fail to identify experts, so they may fail to recognise the epistemic value of minipublics' recommendations. The British Columbia Citizens' Assembly exemplifies this possibility. Chambers complains that shallow rhetoric dominated public discussion of the Assembly's recommended electoral system. Consequently, the proposal was ultimately defeated.⁴³

In light of similar considerations, Lafont concludes that "if we simply insert the conclusion of mini-publics in the political debate in the public sphere without any improvement in the deliberative quality of the latter, then this would make no difference whatsoever".⁴⁴ The idea is that for deliberative minipublics to have any positive impact, we would already need to have improved deliberation in the broader public sphere. Thus, introducing small-scale minipublics in a complementary role does not help mass deliberative democracy avoid the problem of public ignorance.

Lafont's pessimism may be too strong. Minipublics have distinctive features that arguably make them more capable of transmitting knowledge to the broader public than, say, experts. According to MacKenzie and Warren, because of the way in which minipublics are selected, ordinary citizens can more easily recognise their trustworthiness than that of experts. First, random selection means that minipublics, unlike typical groups of experts, are likely to be descriptively representative of the population. So insofar as people are more likely to trust individuals from similar social backgrounds, they are more likely to trust minipublics.⁴⁵ There is some evidence supporting this: Americans, for example, tend to trust the verdicts of citizen juries more than those of judges.⁴⁶ Second, Alex Guerrero argues that because minipublics are selected randomly, they do not rely on electoral campaign financing, and are therefore

⁴² Dryzek (2010, 168-170).

⁴³ Chambers (2009, 331).

⁴⁴ Lafont (2015, 58).

⁴⁵ MacKenzie and Warren (2012, 105-107).

⁴⁶ Rasmussen Reports (2014); Harris Interactive (2008).

less vulnerable to co-optation by powerful lobbies. If so, minipublics may be more visibly trustworthy than elected officials or the experts they appoint.⁴⁷

Admittedly, whether MacKenzie and Warren are correct here is an empirical claim that some might dispute. For my purposes, the main point remains that using minipublics in an advisory capacity is entirely compatible with large-scale deliberative democracy. Hence, if minipublics can serve as trusted information proxies in the way MacKenzie and Warren suggest, deliberative democrats can and should happily embrace their use. Indeed, if this result holds, minipublics enhance the deliberative system's ability to educate ordinary citizens. Thus, they alleviate the worries outlined in Section 3. Understood in a complementary advisory role, then, employing small-scale deliberative initiatives at best helps to solve the ignorance-related problems of large-scale deliberation, and at worst simply reproduces those problems.

Because using minipublics in this complementary role risks reproducing the problems of large-scale deliberation, one might recommend employing minipublics in the more radical way. On this understanding, recall, small lottery-selected deliberative groups are empowered to actually make legislative decisions. Thus, they replace normal sites of mass democratic decision-making, such as legislatures composed of elected officials. In recent years, Alex Guerrero and Claudio López-Guerra have both defended this proposal.⁴⁸ Following Guerrero, let us call this proposal 'lottocracy'. In our context, lottocracy's appeal is straightforward. Since the randomly-selected deliberative group no longer needs to have its recommendations ratified by the broader public, these are no longer hostage to widespread ignorance. Lottocracy apparently circumvents the problem of public ignorance.

An immediate concern is that lottocracy may seem undemocratic. Decision-makers are not selected through elections where citizens have an equal vote. Consequently, citizens are subjected to rule by individuals who are not formally accountable to them.⁴⁹ Lottocrats could respond in two ways. First, like Guerrero, they could argue that rule by randomly-selected deliberative groups *is* democratic in important respects.⁵⁰ Alternatively, they might claim that it is relatively unproblematic that lottocratic arrangements are undemocratic. This might be because (contra my suggestion in

⁴⁷ Guerrero (2014, 164).

⁴⁸ Guerrero (2014); López-Guerra (2014).

⁴⁹ Lafont (2015, 42-54).

⁵⁰ Guerrero (2014, 167).

Chapter 1) they deny that democratic decision-making procedures have intrinsic value from the standpoint of justice. Or, like López-Guerra,lottocrats might think that democracy's intrinsic value is outweighed by the instrumental value of having a political system that overcomes the problem of political ignorance—i.e., that is epistemically capable of producing just outcomes despite widespread ignorance.⁵¹

To sidestep these responses, the rest of this chapter will develop a different line of criticism, which does not appeal to democracy's intrinsic value.⁵² Put differently, this criticism will apply even if you deny that democracy is intrinsically valuable for justice. So far, I have only examined solutions to the problem of public ignorance that are traditionally democratic—where 'traditionally democratic' means that, unlike in lottocracy, ultimate decisional authority lies with the entire demos, each of whom at least has an equal right to vote. However, *because* all the proposals examined so far are traditionally democratic, they involve a bottleneck of inclusive decision-making where ignorance generates difficulties. Ignorant citizens are liable to be swayed by the wrong arguments or testimonies (2.1, 3), to vote for the wrong elites (4.1), and to disregard the recommendations of randomly-selected deliberative minipublics (4.2). If you are unconvinced that democracy is intrinsically valuable, you might therefore wish to remove the democratic bottleneck, and opt for a non-democratic response to public ignorance. The next section will show that this temptation is misleading: the most influential non-democratic strategies for avoiding the problem of public ignorance—including lottocracy—merely postpone this problem.

5. Why the Problem of Public Ignorance Extends to Non-Democratic Views

The present section demonstrates that the problem of public ignorance does not recommend shifting to non-democratic systems: contrary to what one might expect, non-democratic systems merely postpone the problem. Now, one might worry that whether or not this is true depends on the non-democratic system in question. I will focus on the two most prominent non-democratic arrangements that have been

⁵¹ López-Guerra (2014, 26-28, 52-56).

⁵² In Chapter 1, Section 3.1, I suggest that democracy *is* intrinsically valuable for justice. But I avoid relying on this here to sidestep controversies concerning 1) whether lottocracy could meaningfully count as democratic, 2) whether democracy's intrinsic value persists in conditions of political ignorance, and 3) how democracy's intrinsic value compares to the instrumental value of producing good outcomes.

advanced in response to widespread ignorance. I begin by examining epistocracy—rule by knowers—before returning to lottocracy. Starting with epistocracy, which is more obviously non-democratic, will be useful for two reasons. On the one hand, its shortcomings serve to highlight the *prima facie* appeal of lottocracy. On the other, the argumentative strategy developed against epistocracy will help identify how lottocracy also falls short. Although I restrict my focus to specific attempts at realising epistocracy and lottocracy, I will supply reasons for thinking that the problems they encounter do not depend on their institutional specificities.

5.1. Epistocracy and Ignorance

Epistocracy is a political system that distributes decision-making power on the basis of prior knowledge. Though its most famous incarnation is Plato's proffered rule by philosopher kings, Jason Brennan has influentially rearticulated this ideal. Against the background of widespread ignorance, epistocracy's appeal is clear: it proposes to empower those and only those who are knowledgeable. So, if epistemically privileged knowers exist, and if they can be identified and empowered, then epistocracy straightforwardly avoids the problem of public ignorance.

One might doubt that there are privileged knowers in politics. Here, let us bracket such concerns. Instead, the key question for epistocracy is how knowers are to be identified and empowered. How do we *transition* to rule by knowers? Brennan's principal suggestion is that we establish a voter qualification examination. The examination asks questions regarding basic facts that are crucial to making competent political decisions. Only those who perform sufficiently well are then enfranchised.⁵³

However, this recommendation only pushes back the transitional question. Even if there could be an adequate qualification examination, the question becomes how we are to design and implement it. Here, the problem of public ignorance resurfaces. A first option, which Brennan leans towards, is that democratic decision-making should decide what counts as competence.⁵⁴ But why think that the demos—ignorant as it supposedly is—will be competent at designing the examination? The examination tests citizens for knowledge that is crucial to good political decision-making. Hence, *insofar* as democratic

⁵³ Brennan (2016, 211-214).

⁵⁴ *Ibid.*, 224.

decision-making can identify the knowledge requirements that the examination tests for, it is capable of good political decision-making. Putting the same point differently, insofar as democratic decision-making is incompetent at identifying good options—as the problem of public ignorance purports to establish—it is incompetent at identifying the examination’s knowledge requirements. (Notice that this observation equally applies if the demos is supposed to identify competent individuals who then design the test. Insofar as the demos can do so, it can elect competent representatives). The key point is not that it is impossible democratically to design a voter qualification examination. Rather, it is that Brennan’s proposal for implementing epistocracy works only insofar as public ignorance fails to undermine democratic decision-making.⁵⁵

If the democratic route to designing and implementing the voter qualification examination is unpromising, maybe, instead, those who are knowledgeable should somehow forcibly impose the examination. However, this recommendation is question-begging, in the way Chapter 1 discussed.⁵⁶ In political contexts, very often, what is in dispute *just is* who is knowledgeable regarding what we should do. Parties in political disagreements typically take themselves to be right, and their adversaries wrong. Moreover, a characteristic feature of non-ideal conditions characterised by widespread ignorance is that (as Brennan himself suggests⁵⁷) the powerful and the wise seldom coincide.

Putting these two observations together yields an undesirable result. If most competing political actors take themselves to be knowledgeable, and if the powerful and the knowledgeable rarely coincide, we have reason to expect that the question-begging recommendation at hand will in practice not achieve rule by the knowledgeable. Rather, it will achieve rule by whomever the powerful believe—however erroneously—to be knowledgeable. Thus, even on this alternative proposal, public ignorance plagues the transition to epistocracy: when the powerful many are

⁵⁵ Though Brennan considers this objection head-on (2016, 224–226), his answer is unsatisfactory. He challenges the assumption that if you know what the competence requirements are then you are yourself competent, by suggesting that you might know the competence requirements *de dicto* but not *de re*. Simply put, you might know *that rulers should know the basic facts of economics*, without knowing *what the basic facts of economics are*. But if this is what Brennan means when saying that democratic decision-making bodies should decide what counts as competence, it again merely postpones the transitional question. If democratic decision-making only determines the competence requirements *de dicto*, then elaborating the test still requires having someone determine what they are *de re*. But how are *they* chosen? If the answer is ‘democratically’, the problem resurfaces.

⁵⁶ Chapter 1, Section 2.1.c.

⁵⁷ See Brennan’s (2016, 32) assertion that ignorant citizens are far more numerous than knowledgeable citizens.

ignorant and the knowledgeable are few, a policy of having the knowledgeable impose epistocracy seems ill-suited to achieving rule by genuine knowers.

Epistocrats might respond that these worries about transitioning to epistocracy are simply theoretical, and rest upon speculative claims—e.g., about whether or not the powerful many and the knowledgeable coincide. But the theoretical worries I have raised are far from speculative. The real-world track record of voter qualification examinations is terrible. In the American South, Jim Crow laws implementing literacy tests were widely used to disenfranchise black Americans after the Civil War. Nor is the worry merely historical. Voter identification laws are regularly used to keep black Americans from voting.⁵⁸ If voter identification requirements continue to be misused, we may worry that, were they to exist today, voter qualification requirements too might well be misapplied.

Brennan considers this observation, and retorts that it only shows that voter qualification examinations have been poorly implemented in the past. “These tests were administered in bad faith [...] the fact that governments used to hide their racism beneath an epistocratic guise does not show that epistocratic exams are inherently objectionable”.⁵⁹ But this response jars seriously with the broader epistocratic challenge. The relevant question, Brennan continuously stresses, is not ‘is democracy inherently objectionable?’ or ‘is epistocracy inherently objectionable?’ Brennan persistently criticises democratic theorists for defending theories which, *though not inherently bad*, are overly idealised and therefore impracticable in real-world conditions.⁶⁰ Nor is this specific to Brennan. The whole point of the ignorance-based objection to deliberative accounts of democracy is that, though such accounts may be appealing in the ideal, they operate under unrealistic assumptions regarding participants’ cognitive dispositions. So, on pain of inconsistently applying their standards, epistocrats should be asking: given realistic assumptions about what people are like, which political system should we strive for? In this context, the empirical evidence regarding epistocratic arrangements, including voter examinations, is frightful.⁶¹ And this is no accident. It accords with the theoretical argument I offered for thinking that, in non-ideal conditions where power

⁵⁸ E.g., Ingraham (2016).

⁵⁹ Brennan (2016, 223-224).

⁶⁰ Ibid., 69-73, 186-187, 206-207.

⁶¹ See also Popper’s (1995[1945], 145n25) critique of Plato’s Academy, which “was notorious for breeding tyrants”.

and knowledge come apart, we should expect epistocratic experiments to be misapplied.

While I have focused on Brennan's epistocracy, the worries his account faces should generalise. What unites all epistocratic theories is the claim that political decision-making power should be distributed on the basis of prior knowledge. Consequently, they all face the transitional question of how we are to identify and empower knowers. This is precisely the question that replicates democracy's ignorance-related difficulties.

Crucially, the fact that the transitional question plagues epistocracy is far from a superficial criticism. As discussed, epistocrats are explicitly committed to answering this question, insofar as they purport to offer normative guidance for non-ideal circumstances. Fittingly, then, it is the question Plato persistently encountered in *The Republic* and which—by his own admission—he struggled to address.⁶² Thus, epistocracy does not avoid the problem of public ignorance. It simply relocates it to the prior and crucial issue of how we should transition into and maintain epistocracy.

5.2. Lottocracy and Ignorance

This problem might motivate a turn to lottocracy. Unlike epistocracy, lottocracy does not distribute power on the basis of prior knowledge. It distributes power randomly. In turn, once decision-makers have been randomly selected, they undergo a competence-building process where deliberation features centrally. And as 4.2 discussed, there is substantial evidence that deliberative competence-building processes significantly improve the knowledge of randomly-selected citizens.

Thus, López-Guerra claims, by *breeding* rather than *screening for* knowers, lottocracy improves on epistocracy's attempt at avoiding the problem of public ignorance.⁶³ It achieves rule by knowers, thereby keeping power away from ignorant hands. But since there is no need to screen for pre-existing knowers, it apparently circumvents epistocracy's transitional troubles.

Predictably, however, lottocracy only postpones the problem of public ignorance. The crucial premise underpinning lottocracy's appeal is that a competence-building

⁶² Plato (2007[c. 380 BC], 490e-496e).

⁶³ López-Guerra (2014, 25-26).

process can make randomly-selected citizens sufficiently knowledgeable to rule well. Correspondingly, the crucial question is how this competence-building process operates. Either the competence-building process substantively involves interventions by experts, or it does not. If it does not, then, in the present context, it seems unclear why deliberation would make the randomly-selected group more knowledgeable. After all, in Section 3, deliberative democracy was criticised precisely on the grounds that deliberation among ignorant ordinary citizens is allegedly epistemically fruitless. If, however, experts *do* play an essential role—and this is what theorists of minipublics generally suggest⁶⁴—then we have another iteration of the problem epistocracy faced: *transitioning* into lottocracy requires identifying and selecting expert knowers. Just as public ignorance plagued epistocracy with respect to this issue, so it should plague lottocracy.

Guerrero's influential articulation of lottocracy brings this out. Like all lottocrats, he asserts that expert presentations constitute an essential feature of the competence-building process.⁶⁵ Now, in some cases, he claims, it is “relatively uncontroversial whether someone counts as an expert”.⁶⁶ But insofar as this is true, the suggestions Section 3 entertained regarding how deliberative democracy helps eliminate ignorance should also work. Indeed, recall that one of the main deliberative strategies for reducing ignorance involved having citizens seek out and question experts.

In other cases, Guerrero acknowledges that who counts as an expert is more controversial.⁶⁷ How do we identify experts in such cases? If the answer is through democratic decision-making, then Guerrero cannot consistently hold both that this will work well *and* that democratic decision-making will be incompetent at electing representatives. As in the critique of epistocracy, the point is that *if*—as the problem of public ignorance purports to show—democratic decision-making is ineffective at identifying competent and well-motivated representatives, it is unclear why it would be effective at identifying competent and well-motivated experts. So this strategy for identifying knowers functions *only insofar* as public ignorance fails to undermine democratic decision-making.

⁶⁴ Among lottocrats, see Guerrero (2014, 156), López-Guerra (2014, 35-37). Among defenders of advisory minipublics, see Fishkin and Luskin (2005, 288-289), Goodin and Dryzek (2006, 223-224).

⁶⁵ Guerrero (2014, 156).

⁶⁶ *Ibid.*, 161.

⁶⁷ *Ibid.*

Alternatively, if the suggestion is that some competent agency should impose experts through non-democratic means, we have seen when critiquing epistocracy that this solution is objectionably question-begging: in context, who is competent just is what is in dispute. And since the powerful and the knowledgeable often come apart in non-ideal contexts, there are reasons to expect that the most powerful, rather than those who are actually good at identifying experts, will get their way.⁶⁸ Ultimately, Guerrero concedes that “a full defense of the lottocratic alternative will have to do more to specify the details about the [expert] qualification assessment process”.⁶⁹ Thus, he acknowledges that the selection of experts in contexts of ignorance is a problem for lottocracy, without offering a solution.

Because of these difficulties, lottocrats might be tempted to criticise the other horn of the dilemma, and suggest that small-scale deliberation remains epistemically fruitful even without experts. Guerrero, for instance, briefly mentions the possibility of “eliminat[ing] the expert stage”.⁷⁰ The problem with this suggestion is that the evidence supporting the epistemic value of deliberative minipublics overwhelmingly concerns cases involving expert contributions. Deliberative polls explicitly involve interventions by panels of experts. The British Columbia Citizens’ Assembly enlisted experts who taught the assembly about different electoral systems. And the Icelandic Constitutional Assembly was facilitated by experts who supplied templates of what a suitable constitution might look like.

This is not definitively to say that deliberative minipublics cannot build competence without experts. Rather, the point is that if they take this route, lottocrats cannot appeal to the existing empirical data on minipublics. Therefore, they lose a key incentive for embracing lottocracy: one of the central motivations for advocating small-scale deliberation that replaces large-scale deliberative democratic processes was the concrete evidence that small-scale initiatives are epistemically fruitful. Thus, while involving experts replicates the ignorance-related difficulties epistocracy faces, eliminating experts should temper our optimism regarding the epistemic fruitfulness of lottocracy.

⁶⁸ For related worries, see Shapiro (2003, 33).

⁶⁹ Guerrero (2014, 162).

⁷⁰ Ibid., 175.

6. Conclusion

Ignorance regarding political matters threatens to undermine the effectiveness of democratic public deliberation at realising justice. Many have taken this to warrant rejecting democratic theories that promote inclusive deliberation. In resisting this argument, two strategies stand out. The first denies that widespread ignorance impedes public deliberation's ability to promote justice. On this view, inclusive deliberation can eliminate ignorance by pooling knowledge that is distributed across the demos. This strategy has merit. In previous chapters, I explored various respects in which inclusive deliberation contributes to reducing ignorance.⁷¹ The main concern, however, is that inclusive deliberation can only do so when ordinary citizens already meet a certain threshold of knowledge. When they do not, this response may be unsuccessful.

But the present chapter has argued that another strategy is available. Even if ignorance undercuts the justice-promoting powers of deliberative democracy, it is equally problematic for the normative models offered as alternatives. Elite democracy does not avoid the problem, since it relies on ordinary citizens to select elites. Nor is advocating small-scale deliberation within minipublics viable as an alternative to deliberative democracy. Not only does public ignorance pose difficulties for using minipublics in an advisory capacity, but—more importantly—even insofar as there are reasons to think that advisory minipublics help attenuate public ignorance, their use is entirely consistent with deliberative democracy. A more extreme alternative recommends departing from democracy altogether. However, neither of the two most promising non-democratic proposals—rule by empowered minipublics and rule by knowers—truly circumvent the problem. Instead, they simply relocate it to an earlier stage, where we ask how we are to transition to these non-democratic models. So, if widespread ignorance seriously threatens deliberative democracy, then it also seriously threatens the models advanced as replacements. Everyone should be afraid of public ignorance.

Where does this argument leave us? Its positive aim is to refocus our attempts at dealing with public ignorance. First, because ignorance plagues all the competing political models, it does not recommend departing from deliberative theories of

⁷¹ Chapter 1, Section 3.1; Chapter 2, Section 4; Chapter 4, Section 4.

democracy. Since inclusive deliberation is not the source of the problem, our focus should not be on deemphasising it. It should be on tackling public ignorance itself.

Second, given that all normative political models face the problem of public ignorance, inclusive deliberation is better understood as an attempted solution to this problem, an attempt at offsetting widespread ignorance. Now, one sobering conclusion of this chapter is that such deliberation cannot be the only solution to widespread ignorance, for the reason discussed in Section 3: that its effectiveness at reducing ignorance presupposes that some threshold of knowledge has already been met.

Nevertheless, inclusive deliberation may be a fruitful *part* of the solution. The present chapter has encountered some grounds for thinking so. For instance, as discussed in 4.2, there are *prima facie* reasons to expect that using minipublics as one arena within the deliberative system facilitates the transmission of knowledge to the demos. This, recall, is because deliberative minipublics may be easier to identify as trustworthy. The promise of such innovative deliberative practices suggests that despite the difficulties inclusive deliberation faces in countering ignorance, it would be a mistake to jettison it altogether.

In sum, the present chapter should steer deliberative theorists away from the siren songs calling for us to abandon deliberation-centred democracy. Because public ignorance is a problem for all accounts, it supplies no reason to depart from inclusive deliberation. Instead, deliberative theorists should keep their attention directed on a more productive task: to rethink democratic deliberation, both conceptually and institutionally, in a way that makes it a better tool for dispelling ignorance and thereby advancing justice under realistic conditions. If abandoning inclusive deliberation allowed us readily to avoid the problem of public ignorance, we might lose patience with this arduous task. But since alternative models do not avoid this problem, we should instead redouble our efforts.

CONCLUSION

1. Summary of the Argument

I have articulated an account of what norms should govern democratic public deliberation for it to be effective at realising justice in non-ideal conditions. Inclusive public reason-giving, as we have seen, can relate positively to justice in at least two important ways. The first is instrumental: inclusive deliberation contributes to pooling knowledge that is distributed across society, and thereby helps us identify just policies. The second is intrinsic: participating in public deliberation is a way of holding other political actors accountable. Thus, because discursive processes constitute a check against arbitrary uses of power, they are constitutive of non-domination.¹

By way of summary, recall how the discursive norms I have advanced are meant to achieve these justice-promoting functions. Firstly, the requirement that participants in formal deliberation appeal to shared considerations constitutes an important bulwark against domination. By facilitating political contestation, this constraint makes it more difficult for political power to be exercised arbitrarily.

Nevertheless, this norm might seem overly exclusionary. In divided settings, even its weakest formulation excludes some considerations that should have a place in public discourse. To offset these exclusionary implications, I argued that public deliberation should also welcome emotionally-charged narrative, which plays a valuable epistemic role by publicising the experiences of misunderstood social groups.² In particular, because of anger's distinctive phenomenology, anger-infused narratives can play an indispensable role in highlighting overlooked injustices and enhancing listeners' understanding of these injustices.³

However, inviting impassioned personal narratives may seem to be inviting too much. Some citizens' impassioned narratives may (and often do) express highly disrespectful views. When this is the case, one might worry that public deliberation will enact injustice rather than oppose it. To counteract such public hate speech, I argued that democratic public discourse should be governed by robust norms of state-

¹ Chapter 1, Section 3.1.

² Chapter 2.

³ Chapter 4.

sponsored counterspeech aimed at protecting the good standing of targets of hate speech.⁴

Now, this account of public deliberation may seem too demanding in terms of goodwill and of political knowledge. Both problems are real and of vital practical relevance. But philosophical considerations, I have suggested, can go some way towards alleviating them. Firstly, in light of the systemic understanding of deliberation, inclusive deliberation demands less trust and goodwill than critics often suppose. Secondly, there are *prima facie* philosophical reasons to think that trust and goodwill can partly be generated via deliberation. Indeed, philosophical work on trust, hypocrisy, anger, and hate speech provides theoretical reasons to expect that even non-ideal deliberation offers important tools for rebuilding trust and goodwill.⁵

Of course, the second part of this answer depends on empirical considerations as well as philosophical ones. Nonetheless, I have suggested that existing empirical evidence, though incomplete, seems congruent with these theoretical hypotheses. Furthermore, the theoretical hypotheses concerning deliberative trust-building that philosophical work generates should help direct and enrich future empirical research on this topic.

The problem of public ignorance is perhaps the most daunting concern facing deliberative accounts of democracy. In keeping with the last chapter's sobering conclusion, the normative account of deliberation I defend sometimes depends on participants having met a certain threshold of knowledge. Accordingly, public ignorance may undercut deliberation's justice-promoting powers. But public ignorance, I have shown, is just as problematic for alternative normative political models, democratic and non-democratic.⁶

The upshot is this. *Before* the threshold of political knowledge for deliberation to be fruitful is met, public ignorance also plagues alternative normative models. Here, then, deliberative theorists have no reason to abandon inclusive deliberation for an alternative political arrangement. *After* the threshold is met, the justice-promoting functions of inclusive deliberation come into play. In these conditions, the virtues of inclusive deliberation that I have theorised can play a positive part in combatting ignorance and counteracting domination.

⁴ Chapter 3.

⁵ Chapter 5.

⁶ Chapter 6.

So, deliberative theorists should resist the urge to abandon inclusive deliberation because of public ignorance. Instead, going forward, they should focus on more productive endeavours: they should examine how public ignorance might be reduced; and in parallel, they should continue investigating how we might refine the deliberative ideal, conceptually and institutionally, to make it more capable of realising justice despite political ignorance.

2. The Challenge Ahead

My defence of this account of deliberation does not purport to be exhaustive. As I have noted throughout, the account I have offered motivates further, more targeted, lines of investigation, philosophical and empirical. Some pertain to social and moral psychology. While my thesis clarifies the democratic role of emotional speech by exploring angry narratives, the discussion of anger itself invites further research. Under what social conditions, we must ask, is angry testimony most capable of arousing anger in listeners, and thereby performing its distinctive epistemic function? Relatedly, are the kinds of expressions of anger which perform these epistemic functions of the same type as those which, I have suggested, facilitate trust-building?

Another concerns philosophy of language. Authoritative counterspeech can block one of public hate speech's most significant harms—the way it erodes its targets' assurance of their good standing. But philosophers of language have theorised numerous other speech-based harms. Consequently, we must explore how far the template I have offered for countering hate speech can extend to other speech-based harms. While this is partly an empirical task, concerning the expected causal consequences of hate speech when authoritative speakers denounce it, it is also conceptual. Determining whether counterspeech is a fitting response to hate speech requires conceptualising the nature of the harms hate speech is liable to enact.

Lastly, while I have argued for norms of public deliberation, I have only gestured at their institutional manifestations. If counterspeech is needed, how can we empower political agents to perform it systematically and authoritatively? And if angry narrative has a legitimate role to play, how exactly might we alter existing parliamentary and judicial practices to accommodate it? Further social scientific work is needed to

determine precisely what institutional policies would best realise the picture of public deliberation I have sketched.

Investigating these complementary lines of research is a challenging task. But this is no reason not to undertake it. We have seen that attempted ‘shortcuts’ which jettison inclusive deliberation actually reproduce some of its greatest problems. Furthermore, arduous though this task may be, it is also extremely rewarding. By integrating the wide-ranging insights of philosophers and social scientists within democratic theory, it yields a deeper understanding of the democratic ideal, of its relation to justice, and of the place of speech within that ideal. The reward is also political. Where democratic public speech is intimately bound up with exercises of power, and where that speech circulates in increasingly complex ways, we urgently need a theoretical framework that enables us to evaluate and improve upon existing rhetorical practices. The present thesis has outlined such a framework and demonstrated its promise. The task ahead is to fulfil that promise.

BIBLIOGRAPHY

- Abizadeh, Arash. "On the philosophy/rhetoric binaries." *Philosophy & Social Criticism* 33, no. 4 (2007): 445-472.
- Achen, Christopher, and Larry Bartels. *Democracy for Realists*. Princeton, NJ: Princeton University Press, 2016.
- Ahlstrom-Vij, Kristoffer. "Democracy and Moral Conflict." *The Philosophical Quarterly* 61, no. 244 (2011): 666-68.
- Aiken, Nevin. "Learning to Live Together." *International Journal of Transitional Justice* 4 (2010): 166-188.
- Alexander, Larry. "Brettschneider and 'democratic persuasion'." *Jurisprudence* 8, no. 2 (2017): 370-379.
- Allcott, Hunt, and Matthew Gentzkow. "Social Media and Fake News in the 2016 Election." Manuscript, 2017.
- Anderson, Elizabeth. "What Is the Point of Equality?" *Ethics* 109, no. 2 (1999): 287-337.
- . "The Epistemology of Democracy." *Episteme* 3, no. 1-2 (2006): 8-22.
- . *The Imperative of Integration*. Princeton, NJ: Princeton University Press, 2010.
- . "Democracy, Public Policy, and Lay Assessments of Scientific Testimony." *Episteme* 8, no. 2 (2011): 144-64.
- Anderson, Richard. "The Place of the Media in Popular Democracy." *Critical Review* 12, no. 4 (1998): 481-500.
- Arneson, Richard. "Equality and Equal Opportunity for Welfare." *Philosophical Studies* 56 (1989): 77-93.
- . "Luck Egalitarianism and Prioritarianism." *Ethics* 110, no. 2 (2000): 339-49.
- Asser, Martin. "What the Muhammad cartoons portray." BBC, 2010. Available at: http://news.bbc.co.uk/1/hi/world/middle_east/4693292.stm (Accessed 20, May 2016).
- Augustine. "Letter 185: Augustine to Boniface (c.147)." In *Political Writings*, edited by E.M. Atkins and R.J. Dodaro, 173-203. Cambridge: Cambridge University Press, 2001.
- Averill, James. *Anger and Aggression*. New York, NY: Springer, 1982.
- Bächtiger, Andre, Simon Niemeyer, Michael Neblo, Marco Steenbergen, and Jurg Steiner. "Disentangling Diversity in Deliberative Democracy." *Journal of Political Philosophy* 18, no. 1 (2010): 32-63.
- Badano, Gabriele, and Matteo Bonotti. "Rescuing (and Enriching) Public Reason Liberalism's Accessibility Requirement." Manuscript, 2016.
- Baier, Annette. "Trust and Antitrust." *Ethics* 96, no. 2 (1986): 231-260.
- Baker, Edwin. "Harm, Liberty, and Free Speech." *Southern California Law Review* 70 (1996): 979-1020.
- Baldwin, James, Emile Capouya, Lorraine Hansberry, Nat Hentoff, Langston Hughes, and Alfred Kazin. "The Negro in American Culture." *CrossCurrents* 11, no. 3 (1961): 205-24.
- Baldwin, Lewis. "Malcolm X and Martin Luther King, Jr.: What They Thought About

- Each Other.” *Islamic Studies* 25, no. 4 (1986): 395–416.
- Barsade, Sigal. “The Ripple Effect: Emotional Contagion and Its Influence on Group Behaviour.” *Administrative Science Quarterly* 47 (2002): 644–75.
- Bell, Macalester. “Anger, Virtue, and Oppression.” In *Feminist Ethics and Social and Political Philosophy*, edited by Lisa Tessman, 165–83. London: Springer, 2009.
- Bellamy, Richard. *Liberalism and Pluralism*. New York, NY: Routledge, 1999.
- . “Republicanism, Democracy, and Constitutionalism.” In *Republicanism and Political Theory*, edited by Cecile Laborde and John Maynor, 159–89. Oxford: Blackwell, 2008.
- Benhabib, Seyla. “Toward a Deliberative Model of Democratic Legitimacy.” In *Democracy and Difference*, edited by Seyla Benhabib, 67–94. Princeton, NJ: Princeton University Press, 1996.
- . *Another Cosmopolitanism*. Oxford: Oxford University Press, 2006.
- . *Dignity in Adversity*. Malden, MA: Polity Press, 2011.
- Bennett, Stephen, and Linda Bennett. “Out of Sight, Out of Mind: Americans’ Knowledge of Party Control of the House of Representatives, 1960–1984.” *Political Research Quarterly* 46, no. 1 (1993): 67–80.
- Ben-Ze’Ev, Aaron. “Emotions Are Not Mere Judgments.” *Philosophy and Phenomenological Research* 68, no. 2 (2004): 450–57.
- Blairy, Sylvie, Pedro Herrera, and Ursula Hess. “Mimicry and the Judgment of Emotional Facial Expressions.” *Journal of Nonverbal Behavior* 23, no. 1 (1999): 5–41.
- Blake, Michael. “Liberal Foundationalism and Agonistic Democracy.” *Nomos* 46 (2005): 230–43.
- Blow, Charles. “Paul Ryan, Culture and Poverty.” *The New York Times*, 2014. Available at: http://www.nytimes.com/2014/03/22/opinion/blow-paul-ryan-culture-and-poverty.html?_r=1 (Accessed 20 May, 2016).
- Bohman, James. “Deliberative Toleration.” *Political Theory* 31, no. 6 (2003): 757–79.
- . *Democracy Across Borders*. Cambridge, MA: MIT Press, 2007.
- Bohman, James, and Henry Richardson. “Liberalism, Democracy, and ‘Reasons that All Can Accept’.” *Journal of Political Philosophy* 17, no. 3 (2009): 253–274.
- Bonotti, Matteo. “Religion, hate speech, and non-domination.” *Ethnicities* 17, no. 2 (2017): 259–274.
- Borgwald, Kristin. “Women’s Anger, Epistemic Personhood, and Self-Respect.” *Philosophical Studies* 162, no. 1 (2012): 69–76.
- Bourgeois, Patrick, and Ursula Hess. “The Impact of Social Context on Mimicry.” *Biological Psychology* 77 (2008): 343–52.
- Bradshaw, Peter. “Blue Is the Warmest Colour - Review.” *The Guardian*, 2013. Available at: <http://www.theguardian.com/film/2013/nov/21/blue-is-the-warmest-colour-review> (Accessed 15 March, 2016).
- Brady, Michael. *Emotional Insight*. Oxford: Oxford University Press, 2013.
- Brennan, Jason. *Against Democracy*. Princeton, NJ: Princeton University Press, 2016.
- Brettschneider, Corey. *When the State Speaks, What Should It Say?* Princeton, NJ: Princeton University Press, 2012.
- Brisson, Susan. “The Autonomy Defense of Free Speech.” *Ethics* 108, no. 2 (1998): 312–

- Brown, Alexander. *Hate Speech Law: A Philosophical Examination*. London: Routledge, 2015.
- . “What is so special about online (as compared to offline) hate speech?” *Ethnicities* (2017): 1-30.
- Brun, Georg, Ulvi Doguoglu, and Dominique Kuenzle. “Introduction: A New Role for Emotions in Epistemology?” In *Epistemology and Emotions*, edited by Georg Brun, Ulvi Doguoglu, and Dominique Kuenzle, 1–32. Hampshire: Ashgate, 2008.
- Campbell, Bradley, and Jason Manning. “The new millennial ‘morality’: highly sensitive and easily offended.” *Time*, 2015. Available at: <http://time.com/4115439/student-protests-microaggressions/> (Accessed 15 March, 2016).
- Canary, Daniel, and Brian Spitzberg. “Attribution Biases and Associations between Conflict Strategies and Competence Outcomes.” *Communications Monographs* 57 (1990): 139–51.
- Catala, Amandine. “Democracy, Trust, and Epistemic Justice.” *The Monist* 98 (2015): 424–40.
- Chakravarti, Sonali. *Sing the Rage*. Chicago, IL: University of Chicago Press, 2014.
- Chambers, Simone. “Behind Closed Doors: Publicity, Secrecy, and the Quality of Deliberation.” *Journal of Political Philosophy* 12, no. 4 (2004): 389–410.
- . “Rhetoric and the Public Sphere.” *Political Theory* 37, no. 3 (2009): 323–50.
- . “Deliberation and Mass Democracy.” In *Deliberative Systems*, edited by Jane Mansbridge and John Parkinson, 52–71. Cambridge: Cambridge University Press, 2012.
- Christiano, Thomas. “A Democratic Theory of Territory and Some Puzzles about Global Democracy.” *Journal of Social Philosophy* 37, no. 1 (2006): 81–107.
- . “Rational Deliberation among Experts and Citizens.” In *Deliberative Systems*, edited by John Parkinson and Jane Mansbridge, 27–51. Cambridge: Cambridge University Press, 2012.
- Chumley, Cheryl. “Rep. Barbara Lee Slams Bill O’Reilly for ‘Code Words’ That Hide Racism.” *Washington Times*, 2014. Available at: <http://www.washingtontimes.com/news/2014/mar/27/rep-barbara-lee-slams-bill-oreilly-code-words-hide/> (Accessed 20 May, 2016).
- Clift, Tom. “Blue Is the Warmest Colour (Review).” *Moviedex*, 2014. Available at: <http://moviedex.com/reviews/blue-is-the-warmest-colour-review/> (Accessed 15 March, 2016).
- Cogley, Zac. “A Study of Virtuous and Vicious Anger.” In *Virtues and their Vices*, edited by Kevin Timpe and Craig Boyd, 199–224. Oxford: Oxford University Press, 2014.
- Cohen, Gerald Allen. “On the Currency of Egalitarian Justice.” *Ethics* 99, no. 4 (1989): 906–44.
- . *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press, 2008.
- Cohen, Joshua. “Deliberation and Democratic Legitimacy.” In *The Good Polity: Normative Analysis of the State*, edited by Alan Hamlin and Phillip Pettit, 17–34. New York, NY: Blackwell, 1989.

- . “Procedure and Substance in Deliberative Democracy.” In *Democracy and Difference*, edited by Seyla Benhabib, 95–119. Princeton, NJ: Princeton University Press, 1996.
- . “Reflections on Deliberative Democracy.” In *Contemporary Debates in Political Philosophy*, edited by Thomas Christiano and John Christman, 247–63. Malden, MA: Wiley-Blackwell, 2009.
- Conover, Pamela, and Donald Searing. “Studying ‘Every Political Talk’ in the Deliberative System.” *Acta Politica* 40 (2005): 269–83.
- Cooper, Joel, and Russell Fazio. “A New Look at Dissonance Theory.” *Advances in Experimental Social Psychology* 17 (1984): 229–66.
- Coplan, Amy. “Understanding Empathy.” In *Empathy*, edited by Amy Coplan and Peter Goldie, 3–18. Oxford: Oxford University Press, 2011.
- Cramer, Katherine. *The Politics of Resentment*. Chicago: Chicago University Press, 2016.
- Cudd, Ann. *Analyzing Oppression*. Oxford: Oxford University Press, 2006.
- Cyran, Olivier. “‘Charlie Hebdo’, Pas Raciste? Si Vous Le Dites...” *Article 11*, 2013. Available at: <http://www.article11.info/?Charlie-Hebdo-pas-raciste-Si-vous> (Accessed 20 May, 2016).
- Delgado, Richard. “Words That Wound: A Tort Action for Racial Insults, Epithets and Name Calling.” In *Words That Wound*, edited by Mari Matsuda, 89–110. Boulder, CO: Westview Press, 1993.
- Delgado, Richard, and Jean Stefancic. *Understanding Words That Wound*. Boulder, CO: Westview Press, 2004.
- Deonna, Julien, and Fabrice Teroni. *The Emotions: A Philosophical Introduction*. New York, NY: Routledge, 2012.
- Döring, Sabine. “Affective Perception and Rational Motivation.” *Dialectica* 61, no. 3 (2007): 363–94.
- Dougherty, Tom. “Vague Value.” *Philosophy and Phenomenological Research* 89, no. 2 (2014): 352–72.
- Douglass, Frederick. *My Bondage and My Freedom*. Miller, Orton, and Mulligan, 1855.
- Dowding, Keith. “Emotional Appeals in Politics and Deliberation.” *Critical Review of International Social and Political Philosophy* 21, no. 2 (2018): 242–260.
- Dreben, Burt. “On Rawls and Political Liberalism.” In *The Cambridge Companion to Rawls*, edited by Samuel Freeman, 316–46. Cambridge: Cambridge University Press, 2003.
- Dryzek, John. *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. Oxford: Oxford University Press, 2000.
- . *Foundations and Frontiers of Deliberative Governance*. Oxford: Oxford University Press, 2010.
- Dryzek, John, and Christian List. “Social Theory and Deliberative Democracy: A Reconciliation.” *British Journal of Political Science* 33 (2003): 1–28.
- Dworkin, Ronald. “Is There a Right to Pornography?” *Oxford Journal of Legal Studies* 1 (1981): 177–212.
- . “What is Equality? Part 2. Equality of Resources.” *Philosophy & Public Affairs* 10, no. 4 (1981): 283–345.
- . “Foreword.” In *Extreme Speech and Democracy*, edited by Ivan Hare and James

- Weinstein, v–ix. Oxford: Oxford University Press, 2009.
- Elgin, Catherine. “Emotion and Understanding.” In *Epistemology and Emotions*, edited by Georg Brun, Ulvi Doguoglu, and Dominique Kuenzle, 33–50. Hampshire: Ashgate, 2008.
- Ellison, Ralph. *Invisible Man*. New York, NY: Vintage, 1952.
- Elster, Jon. “Strategic Uses of Argument.” In *Barriers to Conflict Resolution*, edited by Kenneth Arrow et al, 236–57. Stanford, CA: Stanford Center on Conflict and Negotiation, 1995.
- . “The Market and the Forum.” In *Deliberative Democracy*, edited by James Bohman and William Rehg, 3–34. Cambridge, MA: MIT Press, 1997.
- . “Deliberation and Constitution Making.” In *Deliberative Democracy*, edited by Jon Elster, 97–122. Cambridge: Cambridge University Press, 1998.
- Enoch, David. “Taking Disagreement Seriously.” *Israel Law Review* 39 (2006): 22–35.
- . “Against Public Reason.” In *Oxford Studies in Political Philosophy, Vol. 1*, edited by David Sobel, Peter Vallentyne, and Steven Wall, 112–40. Oxford: Oxford University Press, 2015.
- Estlund, David. *Democratic Authority*. Oxford: Oxford University Press, 2008.
- Evans, Gareth. *The Varieties of Reference*. Oxford: Oxford University Press, 1982.
- Fabre, Cécile. “A Philosophical Argument for a Bill of Rights.” *British Journal of Political Science* 30, no. 1 (2000): 77–98.
- . “The Dignity of Rights.” *Oxford Journal of Legal Studies* 20, no. 2 (2000): 271–82.
- . *Cosmopolitan Peace*. Oxford: Oxford University Press, 2016.
- Ferguson, Amber. “Elizabeth Warren’s Message at Netroots Nation.” *Huffington Post*, 2015. Available at: http://www.huffingtonpost.com/entry/elizabeth-to-speak-at-netroots-nation-rally_us_55a938f8e4b03f76c5ee2be2 (Accessed 20 August, 2016).
- Festinger, Leon. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press, 1957.
- Finnis, John. “Marriage: A Basic and Exigent Good.” *The Monist* 91, no. 3/4 (2008): 388–406.
- Fishkin, James. *When the People Speak*. Oxford: Oxford University Press, 2009.
- Fishkin, James, and Robert Luskin. “Experimenting with a Democratic Ideal.” *Acta Politica* 40 (2005): 284–98.
- Forst, Rainer. *The Right to Justification: Elements of a Constructivist Theory of Justice*. New York, NY: Columbia University Press, 2012.
- . “Justifying Justification.” In *Justice, Democracy, and the Right to Justification*, edited by Rainer Forst, 169–216. London: Bloomsbury, 2014.
- Fraser, Nancy. *Scales of Justice*. Cambridge: Polity Press, 2009.
- Fricker, Miranda. *Epistemic Injustice*. Oxford: Oxford University Press, 2007.
- . “Epistemic Justice as a Condition of Political Freedom?” *Synthese* 190, no. 7 (2013): 1317–32.
- Friedman, Howard, and Ronald Riggio. “Effect of Individual Differences in Nonverbal Expressiveness on Transmission of Emotion.” *Journal of Nonverbal Behavior* 6, no. 2 (1981): 96–104.
- Friedman, Jeffrey. “Public Ignorance and Democratic Theory.” *Critical Review* 12, no. 4

- (1998): 397–411.
- Friedman, Marilyn. “John Rawls and the Political Coercion of Unreasonable People.” In *The Idea of a Political Liberalism*, edited by Victoria Davion and Clark Wolf, 16–33. Lanham, MD: Rowman & Littlefield, 2000.
- Frye, Marilyn. *The Politics of Reality: Essays in Feminist Theory*. Trumansburg, NY: The Crossing Press, 1983.
- Fung, Archon. “Deliberation Before the Revolution.” *Political Theory* 33, no. 2 (2005): 397–419.
- Garrison, William Lloyd. “Preface.” In *Narrative of the Life of Frederick Douglass*, by Frederick Douglass, iii–xii. Boston: Anti-Slavery Office, 1845.
- Garsten, Bryan. “The Rhetorical Revival in Political Theory.” *Annual Review of Political Science* 14 (2011): 159–80.
- Gaus, Gerald, and Kevin Vallier. “The Roles of Religious Conviction in a Publicly Justified Polity.” *Philosophy & Social Criticism* 35, no. 1–2 (2009): 51–76.
- Gaus, Gerald. *The Order of Public Reason*. Cambridge: Cambridge University Press, 2010.
- . *The Tyranny of the Ideal*. Princeton, NJ: Princeton University Press, 2016.
- Gelber, Katharine. “Reconceptualizing Counterspeech in Hate Speech Policy (with a Focus on Australia).” In *The Content and Context of Hate Speech*, edited by Michael Herz and Peter Molnar, 198–216. Cambridge: Cambridge University Press, 2012.
- Gelber, Katharine, and Luke McNamara. “Evidencing the Harms of Hate Speech.” *Social Identities* 22, no. 3 (2015): 324–41.
- . “The Effects of Civil Hate Speech Laws: Lessons from Australia.” *Law & Society Review* 49, no. 3 (2015): 631–64.
- Geoghegan, Tom. “Who, What, Why: What Is the Quenelle?” *BBC*, 2013. Available at: <http://www.bbc.co.uk/news/blogs-magazine-monitor-25550581> (Accessed 15 May, 2016).
- Gibson, James. “Does Truth Lead to Reconciliation?” *American Journal of Political Science* 48, no. 2 (2004): 201–17.
- Goldberg, Sanford. “On the Epistemic Significance of the Evidence You Should Have Had.” *Episteme* 13, no. 4 (2016): 449–70.
- Goldie, Peter. *The Emotions*. Oxford: Oxford University Press, 2002.
- Goldman, Alvin. “Two Routes to Empathy.” In *Empathy*, edited by Amy Coplan and Peter Goldie, 31–44. Oxford: Oxford University Press, 2011.
- . “Social Epistemology.” *Stanford Encyclopedia of Philosophy*, 2015. Available at: <https://plato.stanford.edu/entries/epistemology-social/> (Accessed 20 May, 2017).
- Goodin, Robert. *Motivating Political Morality*. Oxford: Wiley-Blackwell, 1992.
- Goodin, Robert, and John Dryzek. “Deliberative Impacts: The Macro-Political Uptake of Mini-Publics.” *Politics & Society* 34, no. 2 (2006): 219–44.
- Gray, John. *Two Faces of Liberalism*. Cambridge: Polity Press, 2000.
- Green, Jeffrey. *The Eyes of the People*. Oxford: Oxford University Press, 2010.
- Green, Leslie. “Should Law Improve Morality?” *Criminal Law and Philosophy* 7 (2013): 473–94.
- Guerrero, Alexander. “Against Elections: The Lottocratic Alternative.” *Philosophy &*

- Public Affairs* 42, no. 2 (2014): 135–78.
- Guerrero, Laura. “‘I’m so Mad I Could Scream’.” *Southern Communication Journal* 59 (1994): 125–41.
- Gunn, Paul. “Democracy and Epistocracy.” *Critical Review* 26, no. 1–2 (2014): 59–79.
- Gutmann, Amy, and Dennis Thompson. *Democracy and Disagreement*. Cambridge, MA: Belknap Press, 1996.
- . *Why Deliberative Democracy?* Princeton, NJ: Princeton University Press, 2004.
- Habermas, Jürgen. *Between Facts and Norms*. Cambridge, MA: MIT Press, 1996.
- . “Constitutional Democracy: A Paradoxical Union of Contradictory Principles?” *Political Theory* 29, no. 6 (2001): 766–81.
- . “Religion in the Public Sphere: Cognitive Presuppositions of the ‘Public Use of Reason’ by Religious and Secular Citizens.” In *Between Naturalism and Religion: Philosophical Essays*, 114–48. Cambridge: Polity Press, 2008.
- Hall, Cheryl. “Recognizing the Passion in Deliberation.” *Hypatia* 22, no. 4 (2007): 81–95.
- Halley, Keaton. “How gay marriage harms people.” *Creation.com*, 2017. Available at: <https://creation.com/gay-marriage-harm> (Accessed 20 February, 2018).
- Halpern, Jodi, and Harvey Weinstein. “Rehumanizing the Other: Empathy and Reconciliation.” *Human Rights Quarterly* 26 (2004): 561–83.
- Hardin, Russell. *Trust & Trustworthiness*. New York, NY: Russell Sage Foundation, 2002.
- . *How Do You Know?* Princeton, NJ: Princeton University Press, 2009.
- Harris Interactive. “Just Under Three in Five Americans Believe Juries Can Be Fair and Impartial All or Most of the Time.” *Harris Poll*, 2008. Available at: <http://media.theharrispoll.com/documents/Harris-Interactive-Poll-Research-Just-Under-Three-in-Five-Americans-Believe-Juries-2008-01.pdf> (Accessed 18 May, 2017).
- Harsanyi, David. *The People Have Spoken (and They Are Wrong)*. Washington, DC: Regnery, 2014.
- Hawley, Katherine. “Trust, Distrust, and Commitment.” *Nous* 48, no. 1 (2014): 1–20.
- Heinze, Eric. “Hate Speech and the Normative Foundations of Regulation.” *International Journal of Law in Context* 9, no. 4 (2013): 590–617.
- . *Hate Speech in Democratic Citizenship*. Oxford: Oxford University Press, 2016.
- Hess, Ursula, and Sylvie Blairy. “Facial Mimicry and Emotional Contagion to Dynamic Facial Expressions and Their Influence on Decoding Accuracy.” *International Journal of Psychophysiology* 40 (2001): 129–41.
- Hochschild, Arlie Russell. *Strangers in their Own Land: Anger and Mourning on the American Right*. New York: New Press, 2016.
- Holton, Richard. “Deciding to Trust, Coming to Believe.” *Australasian Journal of Philosophy* 72 (1994): 63–76.
- Honig, Bonnie. “Difference, Dilemmas, and the Politics of Home.” In *Democracy and Difference*, edited by Seyla Benhabib, 257–77. Princeton, NJ: Princeton University Press, 1996.
- Horton, John. “Realism, Liberal Moralism, and a Political Theory of Modus Vivendi.” *European Journal of Political Theory* 9, no. 4 (2010): 431–48.
- Huijter, Marli. “Storytelling to Enrich Democratic Debate: The Dutch Discussion on Embryo Selection for Hereditary Breast Cancer.” *BioSocieties* 4 (2009): 223–238.

- Hume, David. *A Treatise of Human Nature*. Oxford: Oxford University Press, 2009[1738].
- Ingraham, Christopher. "The 'Smoking Gun' Proving North Carolina Republicans Tried to Disenfranchise Black Voters." *The Washington Post*, July 29, 2016. Available at: https://www.washingtonpost.com/news/wonk/wp/2016/07/29/the-smoking-gun-proving-north-carolina-republicans-tried-to-disenfranchise-black-voters/?utm_term=.89848a1fdff6 (Accessed 20 April, 2017).
- Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. "Affect, not Ideology: A Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76, no. 3 (2012): 405-431.
- Jacobs, Lawrence, Fay Lomax Cook, and Michael Carpin. *Talking Together*. Chicago, IL: University of Chicago Press, 2009.
- Jaggar, Alison. "Love and Knowledge: Emotion in Feminist Epistemology." *Inquiry* 32, no. 2 (1989): 151-76.
- Johnson, James. "Arguing for Deliberation." In *Deliberative Democracy*, edited by Jon Elster, 161-84. Cambridge: Cambridge University Press, 1998.
- Jones, Karen. "Trust as an Affective Attitude." *Ethics* 107, no. 1 (1996): 4-25.
- . "Second-Hand Moral Knowledge." *Journal of Philosophy* 94, no. 2 (1999): 55-78.
- Kahn, Rob. "Hate Speech, Democratic Legitimacy, and the Age of Trump." *International and Comparative Law Review* 17, no. 1 (2017): 239-253.
- Kelly, Sean. "Demonstrative Concepts and Experience." *Philosophical Review* 110 (2001): 397-420.
- Kermode, Mark. "Blue Is the Warmest Colour - Review," *The Guardian*, February 27, 2016. Available at: <http://www.theguardian.com/film/2013/nov/24/blue-is-the-warmest-colour-review> (Accessed 13 March, 2016).
- Knops, Andrew. "Debate: Agonism as Deliberation – on Mouffe's Theory of Democracy." *Journal of Political Philosophy* 15, no. 1 (2007): 115-126.
- Kolodny, Niko. "Rule Over None I: What Justifies Democracy?" *Philosophy & Public Affairs* 42, no. 3 (2014): 195-229.
- Krause, Sharon. *Civil Passions*. Princeton, NJ: Princeton University Press, 2008.
- Kukathas, Chandran. *Liberal Archipelago*. Oxford: Oxford University Press, 2003.
- Laborde, Cécile. "Republicanism and Global Justice A Sketch." *European Journal of Political Theory* 9, no. 1 (2010): 48-69.
- Lafont, Cristina. "Deliberation, Participation, and Democratic Legitimacy." *Journal of Political Philosophy* 23, no. 1 (2015): 40-63.
- Landemore, Hélène. *Democratic Reason*. Princeton, NJ: Princeton University Press, 2013.
- . "Yes, We Can (Make It Up on Volume): Answers to Critics." *Critical Review* 26, no. 1-2 (2014): 184-237.
- . "Inclusion Constitution-Making: The Icelandic Experiment." *Journal of Political Philosophy* 23, no. 2 (2015): 166-91.
- Langton, Rae. "Beyond Belief: Pragmatics in Hate Speech and Pornography." In *Speech and Harm*, edited by Ishani Maitra and Mary Kate McGowan, 72-93. Oxford: Oxford University Press, 2012.
- . "Hate Speech and the Epistemology of Justice." *Law and Philosophy*, 2014, 1-9.

- Lazarus, Richard. *Emotion and Adaptation*. Oxford: Oxford University Press, 1991.
- Lenard, Patti. "The Decline of Trust, the Decline of Democracy?" *Critical Review of International Social and Political Philosophy* 8, no. 3 (2005): 363–78.
- . *Trust, Democracy, and Multicultural Challenges*. University Park, PA: Pennsylvania State University Press, 2012.
- Levmore, Saul. "The Internet's Anonymity Problem." In *The Offensive Internet*, edited by Saul Levmore and Martha Nussbaum, 50–67. Cambridge, MA: MIT Press, 2010.
- Lewis, Kristi. "When Leaders Display Emotion." *Journal of Organizational Behavior* 21, no. 2 (2000): 221–34.
- Lincoln, Abraham. "Letter to John A. Gilmer." In *The Collected Works of Abraham Lincoln*, edited by Roy Basler, Vol. 4, 1860.
- List, Christian, and Mathias Koenig-Archibugi. "Can There Be a Global Demos?" *Philosophy & Public Affairs* 38, no. 1 (2010): 76–110.
- Lister, Andrew. "Public Reason and Democracy." *Critical Review of International Social and Political Philosophy* 11, no. 3 (2008): 273–89.
- López-Guerra, Claudio. *Democracy and Disenfranchisement*. Oxford: Oxford University Press, 2014.
- Lorde, Audre. "The Uses of Anger." *Women's Studies Quarterly* 25, no. 1/2 (1997): 278–85.
- Lovett, Frank, and Philip Pettit. "Neorepublicanism: A Normative and Institutional Research Program." *Annual Review of Political Science* 12, no. 1 (2009): 11–29.
- Lukianoff, Greg, and Jonathan Haidt. "The coddling of the American mind." *The Atlantic*, 2015. Available at: <https://www.theatlantic.com/magazine/archive/2015/09/the-coddling-of-the-american-mind/399356/> (Accessed 20 February, 2018).
- Luskin, Robert, Ian O'Flynn, James Fishkin, and David Russell. "Deliberating across Deep Divides." *Political Studies* 62 (2014): 116–35.
- MacKenzie, Michael, and Mark Warren. "Two Trust-Based Uses of Minipublics in Democratic Systems." In *Deliberative Systems*, edited by Jane Mansbridge and John Parkinson, 95–124. Cambridge: Cambridge University Press, 2012.
- MacKinnon, Catharine. *Only Words*. Cambridge, MA: Harvard University Press, 1996.
- Maddison, Sarah. "Relational Transformation and Agonistic Dialogue in Divided Societies." *Political Studies* 63 (2015): 1014–30.
- Maitra, Ishani. "Subordinating Speech." In *Speech and Harm*, edited by Ishani Maitra and Mary Kate McGowan, 94–120. Oxford: Oxford University Press, 2012.
- Maitra, Ishani, and Mary Kate McGowan. "Introduction and Overview." In *Speech and Harm*, edited by Ishani Maitra and Mary Kate McGowan, 1–23. Oxford: Oxford University Press, 2012.
- Malcolm X. *The Ballot or the Bullet*. Cleveland, OH, 1964. Available at: <https://www.youtube.com/watch?v=D9BVENesn6Y> (Accessed 18 August, 2016).
- Manin, Bernard. "On Legitimacy and Political Deliberation." *Political Theory* 15, no. 3 (1987): 338–68.
- Mansbridge, Jane. "Everyday Talk in the Deliberative System". In *Deliberative Politics*, edited by Stephen Macedo, 211–239. Oxford: Oxford University Press, 1999.

- . “Rethinking Representation.” *American Political Science Review* 97, no. 4 (2003): 515–28.
- Mansbridge, Jane, James Bohman, Simone Chambers, David Estlund, Andreas Føllesdal, Archon Fung, Cristina Lafont, Bernard Manin, and José Luis Martí. “The Place of Self-Interest and the Role of Power in Deliberative Democracy.” *Journal of Political Philosophy* 18, no. 1 (2010): 64–100.
- Mansbridge, Jane, James Bohman, Simone Chambers, Archon Fung, John Parkinson, Dennis Thompson, and Mark Warren. “A Systemic Approach to Deliberative Democracy.” In *Deliberative Systems*, edited by John Parkinson and Jane Mansbridge, 1–26. Cambridge: Cambridge University Press, 2012.
- Markovits, Elizabeth. “The Trouble with Being Earnest.” *Journal of Political Philosophy* 14, no. 3 (2006): 249–69.
- Maroh, Julie. “Le Bleu d’Adèle.” *Coeurs-Forêts*, 2013. Available at: <http://www.juliemaroh.com/2013/05/27/le-bleu-dadele/> (Accessed 18 March, 2016).
- Massey, Douglas, and Nancy Denton. *American Apartheid*. Cambridge, MA: MIT Press, 1993.
- Matravers, Derek. *Empathy*. Cambridge: Polity Press, 2017.
- Matsuda, Mari. *Words That Wound*. Boulder, CO: Westview Press, 1993.
- McBride, Cillian. “Reason, Representation, and Participation.” *Res Publica* 13, no. 2 (2007): 171–89.
- McDonnell, Hugh. “How the National Front Changed France.” *Jacobin*, April 22, 2016. Available at: <https://www.jacobinmag.com/2015/11/marine-jean-marie-le-pen-national-front-immigration-elections/> (Accessed 17 May, 2016).
- McGary, Howard. “Forgiveness.” *American Philosophical Association* 26, no. 4 (1989): 343–51.
- Mchangama, Jacob. “The Harm in Hate Speech Laws.” *Policy Review* 176 (2012): 95–102.
- McTernan, Emily. “Microaggressions, Equality, and Social Practices.” *Journal of Political Philosophy* (forthcoming): 1–21.
- Mihai, Mihaela. *Negative Emotions and Transitional Justice*. New York, NY: Columbia University Press, 2016.
- Mill, John Stuart. *On Liberty*. London: Penguin Classics, 2006[1859].
- Mills, Charles. “‘Ideal Theory’ as Ideology.” *Hypatia* 20, no. 3 (2005): 165–84.
- . “White Ignorance.” In *Race and Epistemologies of Ignorance*, edited by Shannon Sullivan and Nancy Tuana, 11–38. Albany, NY: State University of New York Press, 2007.
- Mondon, Aurélien. “The Front National in the Twentieth-Century: From Pariah to Republican Democratic Contender.” *Modern & Contemporary France* 22, no. 3 (2014): 301–20.
- Moore, Alfred. “Anonymity, Pseudonymity, and Deliberation.” *Journal of Political Philosophy* 00, no. 0 (2017): 1–24.
- Morgan-Olsen, Brandon. “Conceptual Exclusion and Public Reason.” *Philosophy of the Social Sciences* 40, no. 2 (2010): 213–43.
- Mouffe, Chantal. “Deliberative Democracy or Agonistic Pluralism?” *Social Research* 66,

- no. 3 (1999): 745–58.
- Murphy, Colleen. *A Moral Theory of Political Reconciliation*. Cambridge: Cambridge University Press, 2010.
- Mutz, Diana. *Hearing the Other Side*. Cambridge: Cambridge University Press, 2006.
- Naurin, Daniel. *Deliberation Behind Closed Doors*. Colchester: ECPR Press, 2007.
- Neblo, Michael, Keavin Esterling, Ryan Kenedy, David Lazer, and Anand Sokhey. “Who Wants to Deliberate - And Why?” *American Political Science Review* 104, no. 3 (2010): 566–83.
- Nielsen, Laura Beth. “Power in Public: Reactions, Responses and Resistance to Offensive Public Speech.” In *Speech and Harm*, edited by Ishani Maitra and Mary Kate McGowan, 148–73. Oxford: Oxford University Press, 2012.
- Nussbaum, Martha. *Upheavals of Thought*. Cambridge: Cambridge University Press, 2001.
- . “Perfectionist Liberalism and Political Liberalism.” *Philosophy and Public Affairs* 39, no. 1 (2011): 3–45.
- . “Transitional Anger.” *Journal of the American Philosophical Association* 1, no. 1 (2015): 41–56.
- . *Anger and Forgiveness*. Oxford: Oxford University Press, 2016.
- . “Beyond Anger.” *Aeon*, 2016. Available at: <https://aeon.co/essays/there-s-no-emotion-we-ought-to-think-harder-about-than-anger> (Accessed 15 February, 2018).
- Oakes, James. *The Radical and the Republican*. New York, NY: W.W. Norton, 2007.
- Okin, Susan Moller. “‘Mistresses of Their Own Destiny’: Group Rights, Gender, and Realistic Rights of Exit.” *Ethics* 112, no. 2 (2002): 205–30.
- O’Neill, John. “The Rhetoric of Deliberation: Some Problems in Kantian Theories of Deliberative Democracy.” *Res Publica* 8 (2002): 249–268.
- Page, Benjamin, and Robert Shapiro. *The Rational Public*. Chicago, IL: Chicago University Press, 1992.
- Parkinson, John. *Deliberation in the Real World*. Oxford: Oxford University Press, 2006.
- Parvin, Phil. “Is Deliberative Democracy Feasible?” *The Monist* 98 (2015): 407–23.
- Peacocke, Christopher. “Does Perception Have a Nonconceptual Content?” *Journal of Philosophy* 98 (2001): 239–64.
- Peoples, Steve, and Emily Swanson. “AP-GfK Poll: Republican voters on Trump: no compassion, no problem.” *AP-GfK*, 2015. Available at: <http://ap-gfkipoll.com/uncategorized/our-latest-poll-findings-45> (Accessed 16 February, 2018).
- Peter, Fabienne. “Democratic Legitimacy and Proceduralist Social Epistemology.” *Politics, Philosophy & Economics* 6, no. 3 (2007): 329–53.
- Pettigrove, Glen. “Meekness and ‘Moral’ Anger.” *Ethics* 122, no. 2 (2012): 341–70.
- Pettit, Philip. “The Cunning of Trust.” *Philosophy & Public Affairs* 24, no. 3 (1995): 202–25.
- . *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press, 1997.
- . “Republicanism: Three Axioms, Four Theorems.” In *Republicanism and Political Theory*, edited by Cécile Laborde and John Maynor, 102–32. Malden, MA: Polity Press, 2008.

- . *On the People's Terms*. Cambridge: Cambridge University Press, 2012.
- Phillips, Anne. "Dealing with Difference: A Politics of Ideas, or a Politics of Presence?" In *Democracy and Difference*, edited by Seyla Benhabib, 139-152. Princeton, NJ: Princeton University Press, 1996.
- Plato. *The Republic*. Translated by Desmond Lee. London: Penguin Classics, 2007[c. 380 BC].
- Polletta, Francesca, and John Lee. "Is Telling Stories Good for Democracy? Rhetoric in Public Deliberation after 9/11." *American Sociological Review* 71 (2006): 699-723.
- Popper, Karl. *The Open Society and Its Enemies, Vol. I*. Routledge: Oxford, 1995[1945].
- Posner, Richard. *Law, Pragmatism, and Democracy*. Cambridge, MA: Harvard University Press, 2003.
- Post, Robert. "Racist Speech, Democracy, and the First Amendment." *William and Mary Law Review* 32 (1990): 267-328.
- . *Constitutional Domains: Democracy, Community, Management*. Harvard University Press. Cambridge, MA, 1995.
- Putnam, Robert. "Bowling Alone: America's declining social capital." *Journal of Democracy* 6, no. 1 (1995): 65-78.
- Quayle, Dan. "The Quayle Quarterly." *Wall Street Journal*. June 8, 1992.
- Quirke, Antonia. "Review - Blue Is the Warmest Colour." *Financial Times*, 2013. Available at: <http://www.ft.com/cms/s/2/86a0b1cc-52af-11e3-8586-00144feabdc0.html#axzz3y4zPKKYZ> (Accessed 14 March, 2016).
- Quong, Jonathan. *Liberalism without Perfection*. Oxford: Oxford University Press, 2011.
- . "On the Idea of Public Reason." In *A Companion to Rawls*, edited by Jon Mandle and David Reidy, 265-280. Oxford: Blackwell, 2013.
- . "Public Reason." *Stanford Encyclopedia of Philosophy*, 2017. Available at: <https://plato.stanford.edu/entries/public-reason/> (Accessed 24 March, 2018).
- Rasmussen Reports. "58% Still Trust A Jury's Verdict More Than A Judge's." *Rasmussen Reports*, 2014. Available at: http://www.rasmussenreports.com/public_content/lifestyle/general_lifestyle/february_2014/58_still_trust_a_jury_s_verdict_more_than_a_judge_s. (Accessed May 19, 2018).
- Rauch, Jonathan. *Kindly Inquisitors*. Chicago, IL: University of Chicago Press, 1993.
- Rawls, John. *Political Liberalism*. New York, NY: Columbia University Press, 1993.
- . "The Idea of Public Reason Revisited." *The University of Chicago Law Review* 64, no. 3 (1997): 765-807.
- . *A Theory of Justice - Revised Edition*. Cambridge, MA: Belknap Press, 1999.
- . *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press, 2001.
- Raz, Joseph. *The Morality of Freedom*. Oxford: Oxford University Press, 1986.
- . "Disagreement in Politics." *American Journal of Jurisprudence* 43, no. 1 (1998): 25-52.
- Reilly, Katie. "Here Are All the Times Donald Trump Insulted Mexico." *Time*, 2016. Available at: <http://time.com/4473972/donald-trump-mexico-meeting-insult/>. (Accessed 16 February, 2018).
- Richards, David. *Free Speech and the Politics of Identity*. Oxford: Oxford University Press,

- 1999.
- Riggio, Ronald. "Power, Persuasion, and Bad Leadership." In *Why Irrational Politics Appeals*, edited by Mari Fitzduff, 71-86. Santa Barbara, CA: ABC-CLIO, 2017.
- Rossi, Enzo, and Matt Sleat. "Realism in Normative Political Theory." *Philosophy Compass* 9 (2014): 689-701.
- Rothman, Lily. "When Spousal Rape First Became a Crime in the U.S." *Time*, 2015. Available at: <http://time.com/3975175/spousal-rape-case-history/> (Accessed 13 March, 2016).
- Rousseau, Jean-Jacques. *The Social Contract*. Translated by G.D.H. Cole. London: Dent, 1973[1762].
- Rowbottom, Jacob. "Government Speech and Public Opinion: Democracy by the Bootstraps." *Journal of Political Philosophy* 25 (2017): 22-47.
- Runciman, David. *Political Hypocrisy*. Princeton, NJ: Princeton University Press, 2008.
- Sacco, Joe. "On Satire - A Response to the Charlie Hebdo Attacks." *The Guardian*, 2015. Available at: <http://www.theguardian.com/world/ng-interactive/2015/jan/09/joe-sacco-on-satire-a-response-to-the-attacks> (Accessed 18 March, 2016).
- Sanders, Lynn. "Against Deliberation." *Political Theory* 25, no. 3 (1997): 347-76.
- Saul, Jennifer. "Dogwhistles, Political Manipulation, and Philosophy of Language." Manuscript, 2017.
- Savransky, Rebecca. "Giuliani: 'Black Lives Matter' Phrase 'Inherently Racist.'" *The Hill*, October 2016. Available at: <http://thehill.com/blogs/blog-briefing-room/news/287157-giuliani-black-lives-matter-inherently-racist> (Accessed 20 February, 2017).
- Scanlon, Thomas. "A Theory of Freedom of Expression." *Philosophy & Public Affairs* 1, no. 2 (1972): 204-26.
- . *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.
- Schaap, Andrew. "Agonism in Divided Societies." *Philosophy & Social Criticism* 32, no. 22 (2006): 255-77.
- Schimmelfenning, Frank. "The Community Trap." *International Organization* 55, no. 1 (2001): 47-80.
- Schulzke, Marcus. "The Social Benefits of Protecting Hate Speech and Exposing Sources of Prejudice." *Res Publica* 22 (2016): 225-42.
- Schumpeter, Joseph. *Capitalism, Socialism, and Democracy*. Abingdon: Routledge, 2010.
- Schwartzman, Micah. "The Sincerity of Public Reason." *Journal of Political Philosophy* 19, no. 4 (2011): 375-98.
- Scott, Dylan, and Alvin Chang. "The Republican tax bill will exacerbate income inequality in America." *Vox*, 2017. Available at: <https://www.vox.com/policy-and-politics/2017/12/2/16720952/senate-tax-bill-inequality> (Accessed 8 February, 2018).
- Scruton, Roger. *Sexual Desire*. London: Phoenix Press, 2001.
- Searing, Donald, Frederick Solt, Pamela Conover, and Ivor Crewe. "Public Discussion in the Deliberative System: Does It Make Better Citizens?" *British Journal of Political Science* 37 (2007): 587-618.
- Seglow, Jonathan. "Hate Speech, Dignity, and Self-Respect." *Ethical Theory and Moral*

- Practice* 19 (2016): 1103–16.
- Sen, Amartya. *The Idea of Justice*. London: Penguin Press, 2009.
- Sénat Vidéo. “Questions Au Gouvernement: David Rachline,” 2015. Available at: <http://videos.senat.fr/video/videos/2015/video30342.html> (Accessed 14 May, 2016).
- Seneca, L.A. “On Anger.” In *Moral Essays*. Translated by John Basore. London: Heinemann, 1928[c.45].
- Sereno, Kenneth, Melinda Welch, and David Braaten. “Interpersonal Conflict.” *Journal of Applied Communication Research* 15 (1987): 128–43.
- Shapiro, Ian. *The State of Democratic Theory*. Princeton, NJ: Princeton University Press, 2003.
- . “On Non-Domination.” *University of Toronto Law Journal* 62, no. 3 (2012): 293–336.
- . “Democracy between Elections.” In *Democracy in the Russian Mirror*, edited by Adam Przeworski, 173–83. Cambridge: Cambridge University Press, 2014.
- . *Politics against Domination*. Cambridge, MA: Belknap Press, 2016.
- Shaw, Rosalind. “Rethinking Truth and Reconciliation Commissions: Lessons from Sierra Leone.” *United States Institute of Peace*. Washington, DC, 2005. Available at: <http://www.usip.org/sites/default/files/resources/sr130.pdf> (Accessed 15 February, 2017).
- Siegmán, Aron Wolfe, Robert Anderson, and Tal Berger. “The Angry Voice.” *Psychosomatic Medicine* 52 (1990): 631–43.
- Silvermint, Daniel. “Rage and Virtuous Resistance.” *Journal of Political Philosophy* (2017): 1–26.
- Simpson, Robert. “Dignity, Harm, and Hate Speech.” *Law and Philosophy* 32 (2013): 701–28.
- Sinaceur, Marwan, and Larissa Tiedens. “Get Mad and Get More than Even.” *Journal of Experimental Social Psychology* 42, no. 3 (2006): 314–22.
- Sleat, Matt. “Coercing non-liberal persons: considerations on a more realistic liberalism.” *European Journal of Political Theory* 12, no. 4 (2013): 347–367.
- Smith, Adam. *The Theory of Moral Sentiments*. Indianapolis: Liberty Classics, 1982[1759].
- Smith, Kimberly. “Storytelling, Sympathy and Moral Judgment in American Abolitionism.” *Journal of Political Philosophy* 6, no. 4 (1998): 356–377.
- Somin, Ilya. “Voter Ignorance and the Democratic Ideal.” *Critical Review* 12, no. 4 (1998): 413–58.
- . “Richard Posner’s Democratic Pragmatism and the Problem of Ignorance.” *Critical Review* 16, no. 1 (2004): 1–22.
- . “Deliberative Democracy and Political Ignorance.” *Critical Review* 22, no. 2–3 (2010): 253–79.
- . *Democracy and Political Ignorance*. Stanford, CA: Stanford University Press, 2013.
- . “Why Political Ignorance Undermines the Wisdom of the Many.” *Critical Review* 26, no. 1–2 (2014): 151–69.
- . “Political ignorance haunts 2016 campaign.” *CNN*, 2016. Available at: <http://edition.cnn.com/2016/05/12/opinions/political-ignorance-somin/> (Accessed 17 January, 2018).

- . “Brexit and political ignorance.” *The Washington Post*, 2016. Available at: https://www.washingtonpost.com/news/volokh-conspiracy/wp/2016/06/14/brexit-and-political-ignorance/?utm_term=.afd7b2a16c93 (Accessed 17 January, 2018).
- Spragens, Thomas. *Reason and Democracy*. Durham, NC: Duke University Press, 1990.
- Srinivasan, Amia. “The Aptness of Anger,” *Journal of Political Philosophy* (2017): 1–29.
- Stanley, Jason. *How Propaganda Works*. Princeton, NJ: Princeton University Press, 2015.
- Steinel, Wolfgang, Gerben Van Kleef, and Fiecke Harinck. “Are You Talking to Me?!” *Journal of Experimental Social Psychology* 44 (2008): 362–69.
- Steiner, Jurg. *The Foundations of Deliberative Democracy*. Cambridge: Cambridge University Press, 2012.
- Steiner Jurg, Maria Clara Jaramillo, Maia Rousiley, and Simona Mameli. *Deliberation across Deeply Divided Societies*. Cambridge: Cambridge University Press, 2017.
- Stromer-Galley, Jennifer. “Measuring Deliberation’s Content: A Coding Scheme.” *Journal of Public Deliberation* 3, no. 2 (2007): 1–35.
- Strossen, Nadine. “Interview with Nadine Strossen.” In *The Content and Context of Hate Speech*, edited by Michael Herz and Peter Molnar, 378–98. Cambridge: Cambridge University Press, 2012.
- Sunstein, Cass. “The Law of Group Polarization.” *Journal of Political Philosophy* 10, no. 2 (2002): 175–95.
- Talisie, Robert. “Deliberativist Responses to Activist Challenges: A Continuation of Young’s Dialectic.” *Philosophy and Social Criticism* 31, no. 4 (2005): 423–44.
- . *Democracy and Moral Conflict*. Cambridge University Press, 2009.
- Tappolet, Christine. “Values and Emotions.” In *Morality and the Emotions*, edited by Carla Bagnoli, 117–31. Oxford: Oxford University Press, 2011.
- Tavris, Carol. *Anger: The Misunderstood Emotion*. New York, NY: Simon & Schuster, 1982.
- Thompson, Dennis. “Deliberative Democratic Theory and Empirical Political Science.” *Annual Review of Political Science* 11 (2008): 497–520.
- Tiedens, Larissa. “Anger and Advancement versus Sadness and Subjugation.” *Journal of Personality & Social Psychology* 80, no. 1 (2001): 86–94.
- Tirrell, Lynne. “Genocidal Language Games.” In *Speech and Harm*, edited by Ishani Maitra and Mary Kate McGowan, 174–221. Oxford: Oxford University Press, 2012.
- . “Transitional Justice in Post-Genocide Rwanda.” In *Theorizing Transitional Justice*, edited by Claudio Corraetti, Nir Eisikovits, and Jack Rotondi, 237–52. Farnham: Ashgate, 2015.
- Tully, James. “Exclusion and Assimilation: Two Forms of Domination in Relation to Freedom.” *Nomos* 45 (2005): 191–229.
- Twohey, Megan. “Rudy Giuliani Lashes Out at Black Lives Matter.” *New York Times*, 2016. Available at: https://www.nytimes.com/2016/07/11/us/politics/rudy-giuliani-black-lives-matter.html?_r=0 (Accessed 25 February, 2017).
- Tye, Michael. “Nonconceptual Content: Richness and Fineness of Grain.” In *Perceptual Experience*, edited by Tamar Gendler and John Hawthorne, 504–30. Oxford: Oxford University Press, 2006.

- UN. International Convention on the Elimination of All Forms of Racial Discrimination (1965).
- Urbinati, Nadia. "Representation as Advocacy." *Political Theory* 28, no. 6 (2000): 758–786.
- Usborne, Simon. "Katie Hopkins has just written a piece so hateful that it might give Hitler pause—why was it published?" *Independent*, 2015. Available at: <http://www.independent.co.uk/voices/katie-hopkins-when-is-enough-enough-10186490.html> (Accessed 15 February, 2018).
- Valentini, Laura. "On the Apparent Paradox of Ideal Theory." *Journal of Political Philosophy* 17, no. 3 (2009): 332–55.
- . "Justice, Disagreement and Democracy." *American Political Science Association* 12, no. 4 (2013): 177–99.
- Vallier, Kevin. "Against Public Reason Liberalism's Accessibility Requirement." *Journal of Moral Philosophy* 8 (2011): 366–89.
- . "Public Justification versus Public Deliberation: The Case for Divorce." *Canadian Journal of Philosophy* 45, no. 2 (2015): 139–58.
- . "In Defence of Intelligible Reasons in Public Justification." *The Philosophical Quarterly* 66, no. 264 (2016): 596–616.
- Van Beest, Ilja, Gerben Van Kleef, and Eric Van Dijk. "Get Angry, Get out." *Journal of Experimental Social Psychology* 44 (2008): 993–1002.
- Waldron, Jeremy. "A Rights-Based Critique of Constitutional Rights." *Oxford Journal of Legal Studies* 13, no. 1 (1993): 18–51.
- . *Law and Disagreement*. Oxford: Oxford University Press, 1999.
- . "The Core of the Case Against Judicial Review." *The Yale Law Journal* 115, no. 6 (2006): 1346–1406.
- . "Vagueness and the Guidance of Action." In *Philosophical Foundations of Language in the Law*, edited by Andrei Marmor and Scott Soames, 58–82. Oxford: Oxford University Press, 2010.
- . *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press, 2012.
- Walker, Shaun. "Vladimir Putin: Gay People at Winter Olympics Must 'Leave Children Alone.'" *The Guardian*, 2014. Available at: <http://www.theguardian.com/world/2014/jan/17/vladimir-putin-gay-winter-olympics-children> (Accessed 14 March, 2016).
- Walzer, Michael. "Philosophy and Democracy." *Political Theory* 9, no. 3 (1981): 379–99.
- Warren, Mark. "Institutionalizing Deliberative Democracy." In *Deliberation, Participation and Democracy*, edited by Shawn Rosenberg, 272–88. New York, NY: Palgrave Macmillan, 2007.
- Weigel, David. "Three words that Republicans wrestle with." *Washington Post*, 2016. Available at: https://www.washingtonpost.com/politics/three-words-that-republicans-wrestle-with-black-lives-matter/2016/07/12/f5a9dfdc-4878-11e6-90a8-fb84201e0645_story.html?utm_term=.0bd766ba0ceb (Accessed 20 February, 2017).
- Weinstock, Daniel. "Compromise, pluralism, and deliberation." *Critical Review of International Social and Political Philosophy* 20, no. 5 (2017): 636–655.
- Wendt, Fabian. "Peace beyond Compromise." *Critical Review of International Social and*

- Political Philosophy* 16, no. 4 (2013): 573–93.
- . “The Moral Standing of Modus Vivendi Arrangements.” *Public Affairs Quarterly* 30, no. 4 (2016): 351–70.
- Wesolowska, Elzbieta. “Social Processes of Antagonism and Synergy in Deliberating Groups.” *Swiss Political Science Review* 13, no. 4 (2007): 663–680.
- West, Robin. “Liberty, Equality, and State Responsibilities.” *Brooklyn Law Review* 79, no. 3 (2014): 1031–45.
- Williams, Melissa. “The Uneasy Alliance of Group Representation and Deliberative Democracy.” In *Citizenship in Diverse Societies*, edited by Will Kymlicka and Wayne Norman, 124–52. Oxford: Oxford University Press, 2000.
- Willsher, Kim. “Abandoned’ French Working Class Ready to Punish Left’s Neglect by Voting for Far Right.” *The Guardian*, 2015. Available at: <http://www.theguardian.com/world/2015/mar/22/alienated-french-working-class-vote-far-right-claims-analyst> (Accessed 18 May, 2016).
- Wubben, Maarten, David De Cremer, and Eric Van Dijk. “How Emotion Communication Guides Reciprocity.” *Journal of Experimental Social Psychology* 45 (2009): 987–90.
- Yong, Caleb. “Does Freedom of Speech Include Hate Speech?” *Res Publica* 17 (2011): 385–403.
- Young, Iris Marion. *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press, 1990.
- . “Communication and the Other: Beyond Deliberative Democracy.” In *Democracy and Difference*, edited by Seyla Benhabib, 120–36. Princeton, NJ: Princeton University Press, 1996.
- . *Inclusion and Democracy*. Oxford: Oxford University Press, 2000.
- . “Activist Challenges to Deliberative Democracy.” *Political Theory* 29, no. 5 (2001): 670–90.